Word prediction

Dataset:

The word prediction model is designed to generate accurate predictions based on a dataset consisting of 4 pages from Wikipedia. The dataset is preprocessed to remove HTML tags, URLs, punctuation, white spaces, and digits. Lemmatization and tokenization techniques are then applied to transform the text into a sequence of words.

Model:

The model architecture consists of an embedding layer followed by 4 simple RNN layers, each containing 50 neurons. The model is designed to learn patterns and dependencies within the input sequence to predict the next word. The final layer of the model is a softmax layer with a number of units equal to the total number of words.

Accuracy:

the model achieves an accuracy of almost 100%, indicating its ability to accurately predict the next word given the previous sequence.

```
Epoch 73/80
36/36 [==============================] - 11s 305ms/step - loss: 0.2537 - accuracy: 0.9991
Epoch 74/80
36/36 [==============================] - 11s 304ms/step - loss: 0.2443 - accuracy: 0.9991
Epoch 75/80
36/36 [==============================] - 11s 299ms/step - loss: 0.2332 - accuracy: 0.9991
Epoch 76/80
36/36 [==============================] - 11s 307ms/step - loss: 0.2280 - accuracy: 0.9982
Epoch 77/80
36/36 [==============================] - 10s 281ms/step - loss: 0.2163 - accuracy: 0.9982
Epoch 78/80
36/36 [==============================] - 10s 282ms/step - loss: 0.2143 - accuracy: 0.9973
Epoch 79/80
36/36 [==============================] - 10s 277ms/step - loss: 0.2010 - accuracy: 0.9991
Epoch 80/80
36/36 [==============================] - 10s 285ms/step - loss: 0.1901 - accuracy: 0.9991
<keras.src.callbacks.History at 0x7ed0dc2b9720>
```

Test :

```
[ ]
seed_text = "I love"
next_words = 2
generated_text = generate_next_word(seed_text, next_words)
print(generated_text)
```

```
1/1 [==============================] - 0s 50ms/step
1/1 [==============================] - 0s 48ms/step
I love football family
```

```
seed_text = "football "
next_words = 3
generated_text = generate_next_word(seed_text, next_words)
print(generated_text)
```

```
1/1 [==============================] - 0s 51ms/step
1/1 [==============================] - 0s 49ms/step
1/1 [==============================] - 0s 49ms/step
football  family team sports
```

conclusion:

the word prediction model demonstrates good performance in generating accurate predictions for the next word in a given sequence. It effectively learns the patterns and dependencies in the input text and provides highly accurate predictions.