**APPROACH FOR THE STATEMENT 1:**

1. **Load and preprocess the data:** The code starts by loading a dataset containing spam and non-spam emails. The encoding of the data is not set so we have to set the encoding otherwise parsing error is coming .The dataset is preprocessed by tokenizing the text
2. **Tokenize and pad sequences:** The text data is tokenized using a tokenizer and converted to sequences of integers. The sequences are then padded to a fixed length to enable them to be fed into a neural network.
3. **Create a neural network model:** The code defines a neural network model with an embedding layer, followed by a flatten layer. The embedding layer learns a dense representation of the input text data, which is then flattened and fed into multiple dense layers for classification.
4. **Train the model:** The model is trained on the preprocessed training data using the binary cross-entropy loss function and the Adam optimizer. **Early stopping is used to prevent overfitting to the training data.**
5. **Evaluate the model:** Finally, the model is evaluated on the preprocessed testing data, and the accuracy of the model is calculated.

**FINAL TEST ACCURACY IS : 98.39%**

**ANSWER TO THE GIVEN QUESTIONS:**

1. Word embedding methods are a good choice for text classification tasks like email spam classification because they provide a way to represent words in a dense, low-dimensional vector space that captures their semantic meaning.Word embeddings can also help to reduce the sparsity of high-dimensional input data, which can lead to better generalization of the model.**( just like PCA and ATTENTION FUNCTION….THEY ALL ARE KIND OF SAME CONVERTING R^d into R^l such that l<<<d)**
2. **The neural network model is a sequential model in Keras, with the following layers:**
3. **Input layer:** The input layer receives the preprocessed text data in the form of integer sequences with fixed length.
4. **Embedding layer:** The embedding layer learns a dense representation of the input text data by mapping each word index to a dense vector. The size of the output vectors is specified by the "embedding\_dim" parameter.
5. **Flatten layer:** The flatten layer is used to convert the output of the embedding layer from a 3D tensor to a 2D tensor, which can be fed into the dense layers.
6. **Output layer:** The output layer is a single neuron with sigmoid activation function that produces a binary classification output indicating whether the email is spam or not.

**The model is trained using binary cross-entropy loss function and the Adam optimizer, and early stopping is used to prevent overfitting to the training data. The model is evaluated using the accuracy metric on the preprocessed testing data.**

1. FINAL TEST ACCURACY IS : 98.39%

**THIS IS THE APPROACH WHICH I APPLIED IF THERE IS ANY SUGGESTIONS OR FEEDBACK REGARDING THIS THEN PLEASE TELL ME**