**Deep learning Based Text Classification.**

**Abstract:**

Text classification deals with unstructured / raw or unknown data and categorizes them based on similarity observed between these data. In this Project we are going to develop a model which will analyze raw data and train the model to categorize the data based on the features we can find using a Neural Network.Deep Learning is based on the concept of Artificial neural network which tries to replicate on how neurons work in our brain. In this we have multiple steps which processes data to progressively extract higher level features from the data.There are various Neural Network Mechanisms or models which have been used for Text Classification such as CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory) Models.Text Classification was previously done with Machine Learning algorithms like Logistic Regression, Random Forest and K-Nearest Neighbors but were outperformed with various Deep Learning Algorithms.

**Literature Overview:**

Transformers For Text Classification

<https://blog.paperspace.com/transformers-text-classification/>

Text Classification Using BERT & Tensorflow

<https://www.youtube.com/watch?v=hOCDJyZ6quA>

Text Classification using Deep Learning: A Comprehensive Review

<https://arxiv.org/pdf/2004.03705.pdf>

Attention is All You Need. – Research Article on transformers architecture.

<https://arxiv.org/pdf/1706.03762.pdf?ref=blog.paperspace.com>

Bibliometric analysis and future research for text classification using deep learning

<https://www.researchgate.net/publication/373208376_Text_classification_using_deep_learning_techniques_a_bibliometric_analysis_and_future_research_directions>

Positional Encoding in Transformers.

<https://machinelearningmastery.com/a-gentle-introduction-to-positional-encoding-in-transformer-models-part-1/>

**Introduction:**

Why do we need Text Classification?

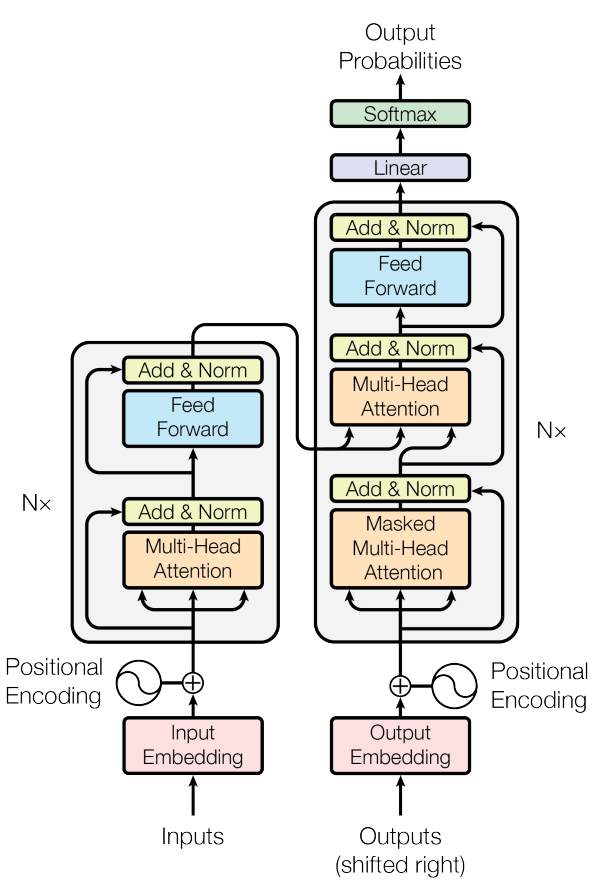
Customers and buyers often check on reviews before purchasing a product and analyze on how good a product is by factors of price, quantity purchased, stars and reviews placed on them.

Although it is easy to read and classify features such as price and quantity, analyzing text is different. Humans are cognizant of the knowledge they have adapted to a specific language, they are able to mostly understand the hidden intent, sarcasm or other critical aspects of speech notions which is not present in machines. Machine Learning (ML) find it hard to decipher the true intent and meaning of a sentence due to its inherent complexity to understand something which is not in a binary form.

Text classification can be described as a machine learning technique to classify the type of text into a particular category. These categories differ based on task they perform such as sentiment analysis, topic labeling, spam detection and intent detection.

As data that available in the natural world is mostly unstructured, it becomes crucial to find a way to manage these datasets to form a desirable structure to analyze these problems.

Purpose Text classification is a widely accepted and adopted technique in organizations to mine and analyze unstructured and semi-structured data. With advancement of technological computing, deep learning has become more popular among academicians and professionals to perform mining and analytical operations. In this work, the authors study the research carried out in field of text classification using deep learning techniques to identify gaps and opportunities for doing research. Design/methodology/approach. The authors adopted bibliometric-based approach in conjunction with visualization techniques to uncover new insights and findings. The authors collected data of two decades from Scopus global database to perform this study. The authors discuss business applications of deep learning techniques for text classification. Findings The study provides overview of various publication sources in field of text classification and deep learning together. The study also presents list of prominent authors and their countries working in this field. The authors also presented list of most cited articles based on citations and country of research. Various visualization techniques such as word cloud, network diagram and thematic map were used to identify collaboration network. Originality/value. The study performed in this paper helped to understand research gaps that is original contribution to body of literature. To best of the authors' knowledge, in-depth study in the field of text classification and deep learning has not been performed in detail. The study provides high value to scholars and professionals by providing them opportunities of research in this area.



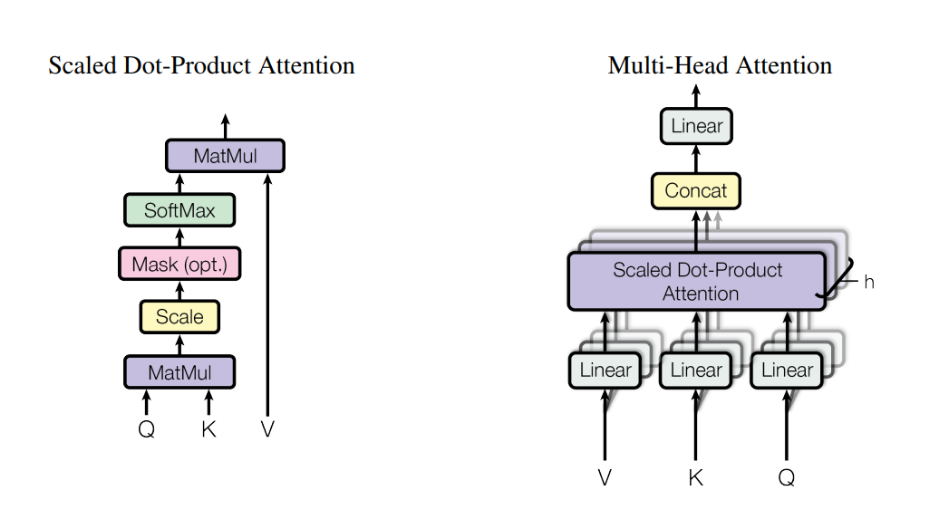
**Transformer Architecture:**

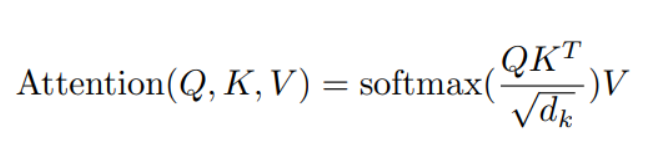
This Image represents the model architecture for Transformers which most competitive neural sequence transduction models use an encoder-decoder structure. The left side of the transformers displays the encoding architecture where we consist of the inputs, input embeddings, positional encodings and a block containing some neural network components. This block is an integral part aspect of understanding on how neural networks work.

The multi head attention in the encoding block takes in 3 inputs being Queries, Keys and Values followed by Concatenation or Addition then Normalization and passed through a Feed Forward Network and then reform the Addition and Normalization.

The embedding input in the form of Queries, Keys and Values is passed through a linear layer, let us say the embeddings have a dimensionality of 64 and we make use of four splits for each layer, then each of the layers passed through will have four blocks of 16-dimensional data embedded in them.

Then the data is passed through a scaled dot-product attention.





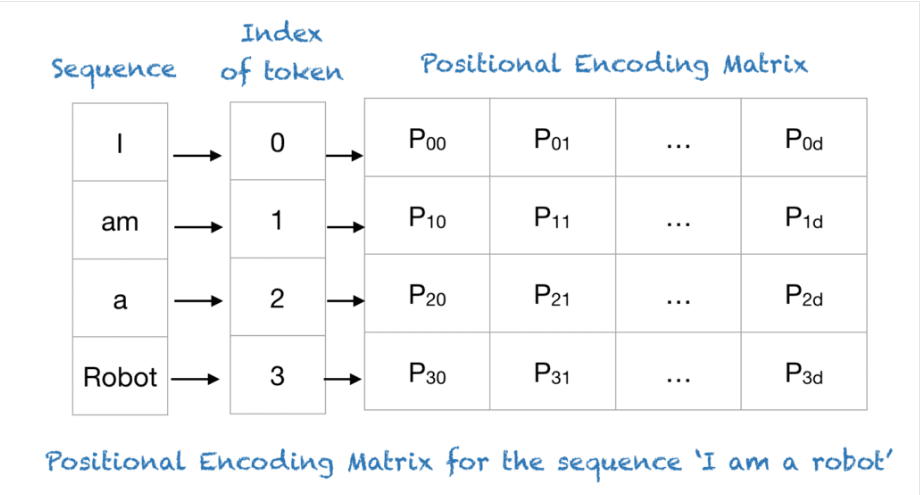
The formula above represents the clear performance of the model which is equal to the SoftMax of the product of queries and transpose of keys divided by the scaling factor (represented by the square root of embedding size) and multiplied with the vectors, then they are concatenated so that they are of the same size as that of the embedding, and then is passed through a linear layer.

Finally, the right-hand side of the transformers network share similarities with the encoding block. They have a couple of transformer blocks, one with masked multi-head attention to ensure that each consecutive step only has the knowledge of the previous step. The model can have multiple encoders and decoders blocks depending on the task.

**Positional Encoding:**

When dealing with languages, order and positions of words in a sentence really matter. The meaning can change with how we can arrange the words in a sentence. In neural networks models they have an inbuilt mechanism to deal with order of sequences. However, in the transformer model we do not use recurrence or convolution and treat each data item independent of each other. Hence a positional Encoding is added to the model explicitly to retain information regarding the order of words in a sequence.

Positional Encoding describes the location or position of an entity in a sequence such that each position is assigned a unique representation. Although indices can be used for sentences, they get large in magnitude as the indices grow in long sequences. If we decide to normalize values to 0 and 1, it can create problems for variables length sequences as they would be normalized differently.



Transformers use a smart positional encoding scheme, with each position/index being mapped to a vector. Hence the output of the positional encoding is a matrix, where each row represents an encoded object of the sequence summed