Advantage of using subword embeddings is that it can perform better in noisy user-generated text like those found in online media.

Future research possibilities of BNLP are: Reliable Resource Design, Accessible Resource, Reliable and Reproducible Benchmark, Strengthen Monolingual Pretrained Transformer Models, Technology Transfer, Broaden the Horizon

Support Vector Machine (SVM) is a simple supervised machine algorithm used for classification and regression purposes.

Advantages of SVM are:

1. It works relatively well when there is a clear margin of separation between classes.
2. It is more effective in high dimensional spaces.
3. It is effective in cases where the number of dimensions is greater than the number of samples.
4. It is relatively memory efficient

Conditional Random Fields is a discriminative undirected probabilistic graphical model, a sort of Markov random field.

LSTM is a type of recurrent neural network but is better than traditional recurrent neural networks in terms of memory. A Bidirectional LSTM, or biLSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction.

The GRU is like a long short-term memory (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate. A Bidirectional GRU, or BiGRU, is a sequence processing model that consists of two GRUs. one taking the input in a forward direction, and the other in a backwards direction. It is a bidirectional recurrent neural network with only the input and forget gates.

Character level embedding uses one-dimensional convolutional neural network (1D-CNN) to find numeric representation of words by looking at their character-level compositions.

The Hugging Face transformers package is an immensely popular Python library providing pretrained models that are extraordinarily useful for a variety of natural language processing (NLP) tasks

BERT, RoBERTa, ALBERT

Bangla-Electra is a method for self-supervised language representation learning.

A deep belief network is a generative graphical model, or alternatively a class of deep neural network, composed of multiple layers of latent variables, with connections between the layers but not between units within each layer

Subword embedding performs better in noisy user-generated text

The big premise of the Transfomer-based models is that they can be pre-trained on huge amounts of unlabeled data (such as all of Wikipedia or a book corpus), and later fine-tuned to a specific task (e.g., question-answering) using just a small amount of labeled, domain-specific data. To achieve high accuracy, those models feature millions (and, at times, billions) of parameters, and require long and expensive training.

XLM-RoBERTa model performs best for emotion classification.

Due to morphological complexity, and highly inflected words, problem occurs to translate rare words in transfer based models

It creates difficulties for the model to predict proper words for the target sequence

No, having additional data does not help to improve the model performance for Parts of Speech tagging.

TERRIER model is used for stemming.

Intuitively, Adam optimization algorithm is a combination of the ‘gradient descent with momentum’ algorithm and the ‘RMSP’ algorithm.

The number of tokens can go up when the title and the subtitle are associated with a person entity

Word counts are weighted similarly so that words from low-resource languages are represented adequately in terms of vocabulary.

Prefix, suffix, the word in context, abbreviation, and numbers are the orthographic features.