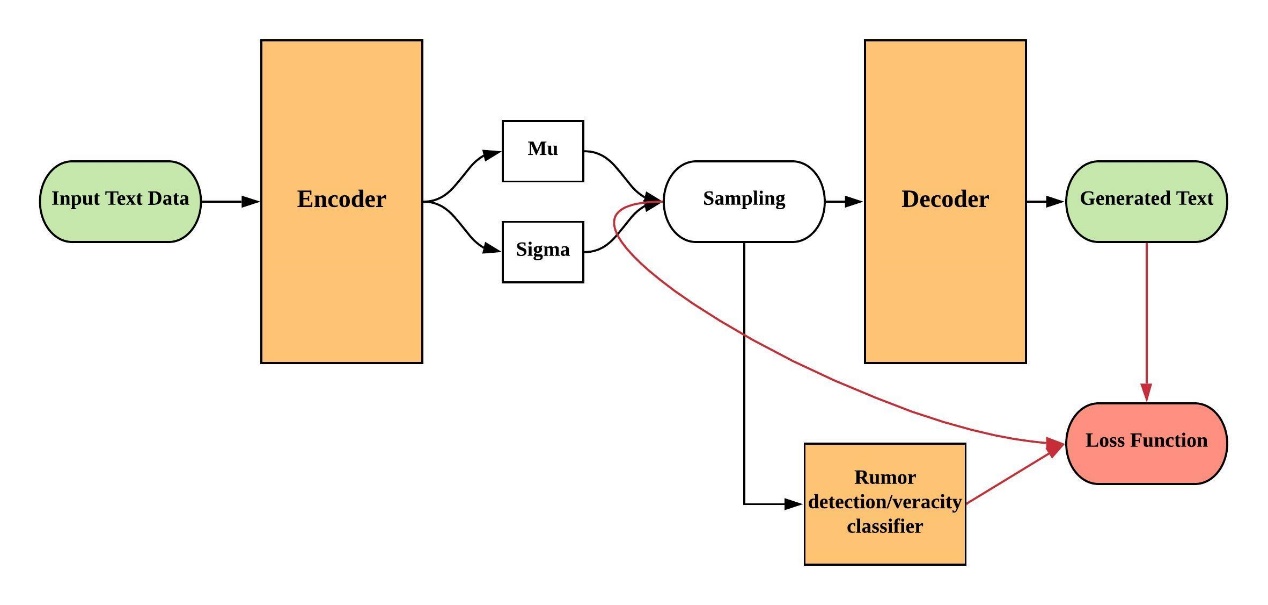
Software Challenges

There are two main challenges in software part. The first is that we plan to build the total framework of the variational autoencoder-aided multi-task rumor classifier. To make things more challenging, we are required to implement the functionality twice with Python and C++.

There are many mature frameworks in Python to implement VAE and LSTM like TensorFlow and Pytorch, so we can just read the documents and accomplish the assignment with fully grasp of the algorithm. Challenges will lie in the process of trading off parameters to get a best performance and faster speed.

We also plan to conduct the framework of the algorithm with C++ to get a better understanding of the techniques and algorithms. The general frame is shown in the flowchart.



The goal is to encapsulate classes and design the structure and APIs in a decent way. In the classes of encoder, decoder and rumor classifiers, it is better to use the Pytorch C++ frontend to implement LSTM and BiLSTM. The green part is input and output text messages, of which input is randomly chosen from the whole training set to realize the stochastic gradient descent. The loss function is the weighted sum of KL divergence, l2 distance between original and generated text, and the accuracy of the rumor classifier. We need to build a connection between the modules.

The second challenge is to use web crawler to get the large amount of data we need on Twitter, Reddit, and other social media. Then we may use filters to get a features dataset. The challenge is that we never use a web crawler, and the filter may be difficult to design in consideration of the data size. This part will be done in Python.