AI NTPU

Dr. Chih-Chien Wang
Dr. Min-Yuh Day
Mr. Wei-Jin Gao
Mr. Yen-Cheng Chiu
Ms. Chun-Lian Wu

National Taipei University
Tamkang University
Taipei, Taiwan

Overview

Retrieval based Method — Solr search engine + Similarity

Generative Model — Short Text Generation; Emotion Classification model

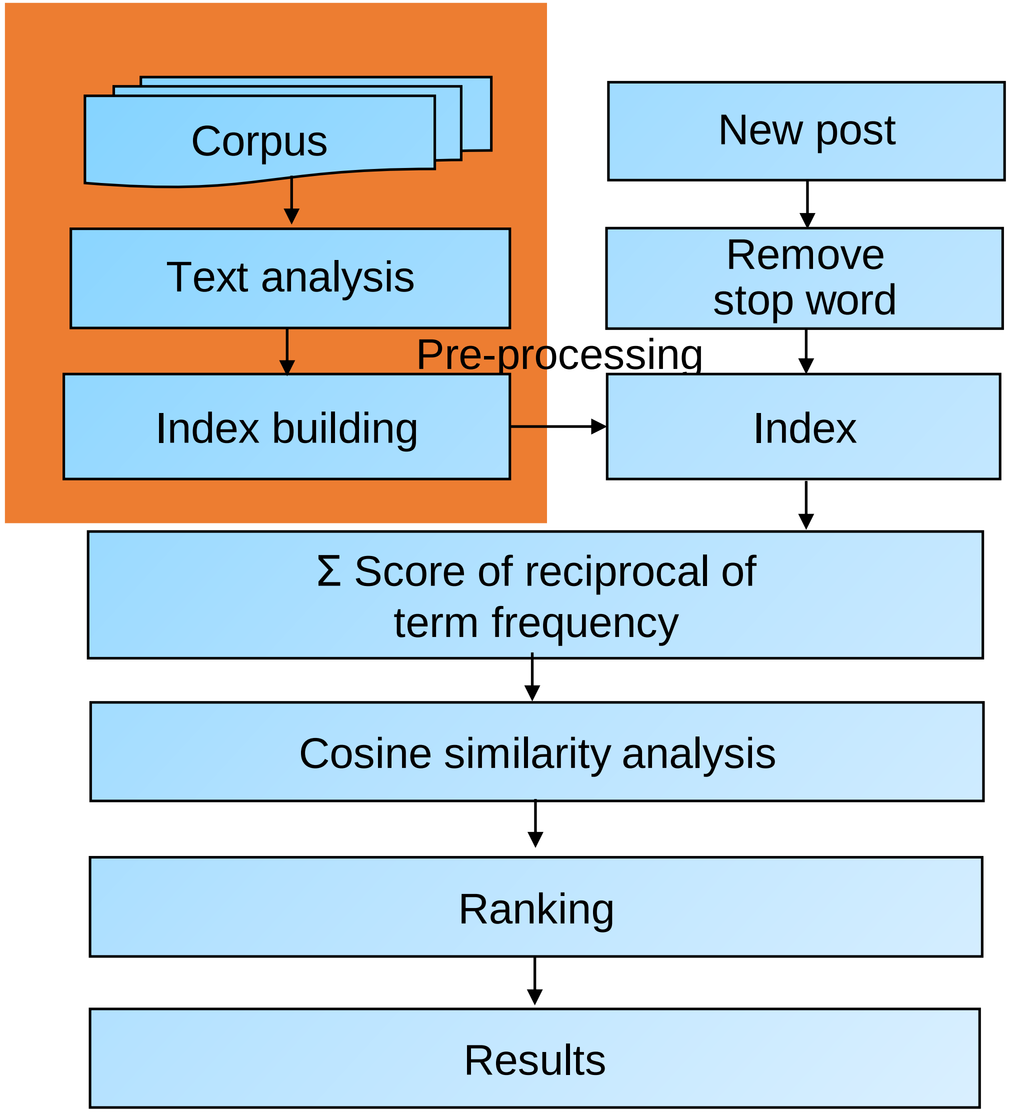
Generative Model + General Purpose Response — Generation Purpose Response

Retrieval Based

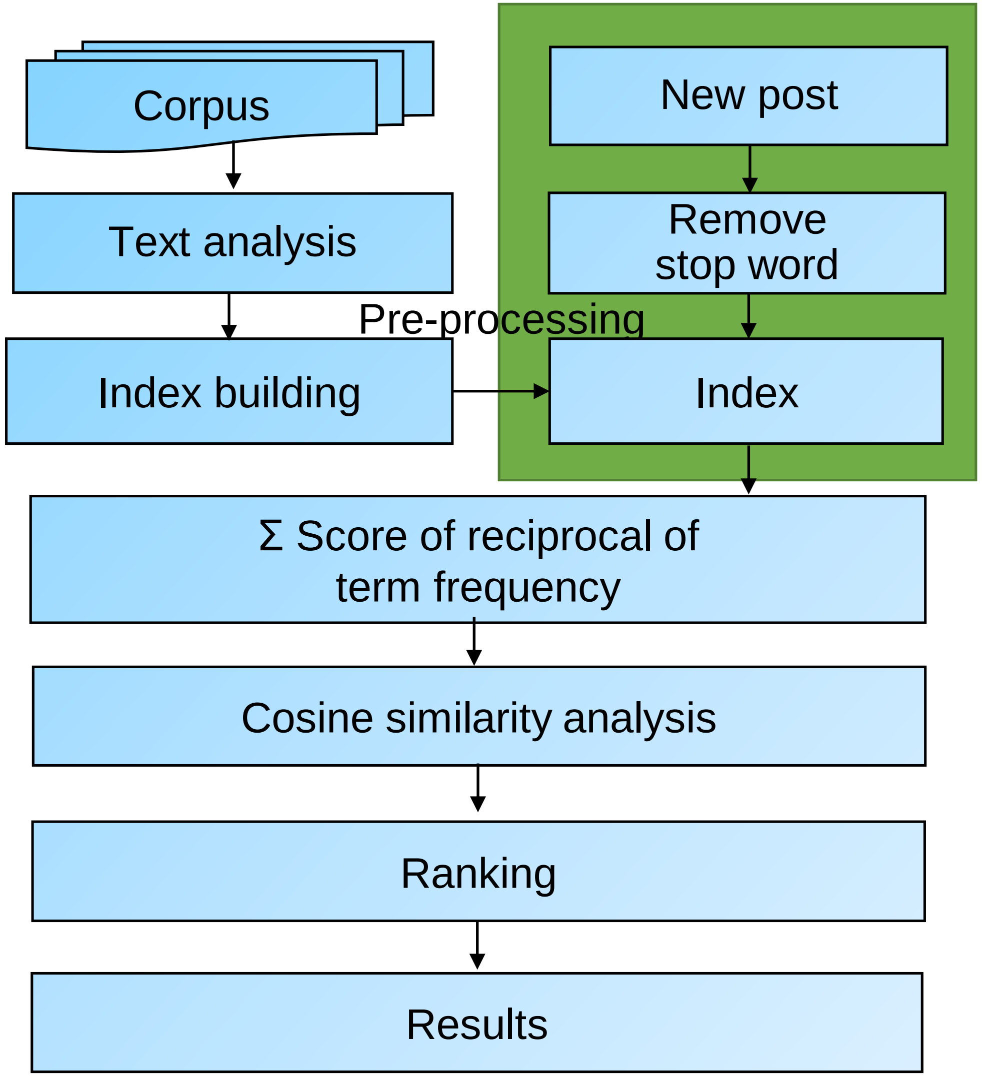
Search responses from corpus.

Overview of Retrieval-based Method

- We used Solr to index the corpus.
- Before indexing it, we perform word segmentation, text analysis, and remove stop words.
- Then, we complete the Solr index building.

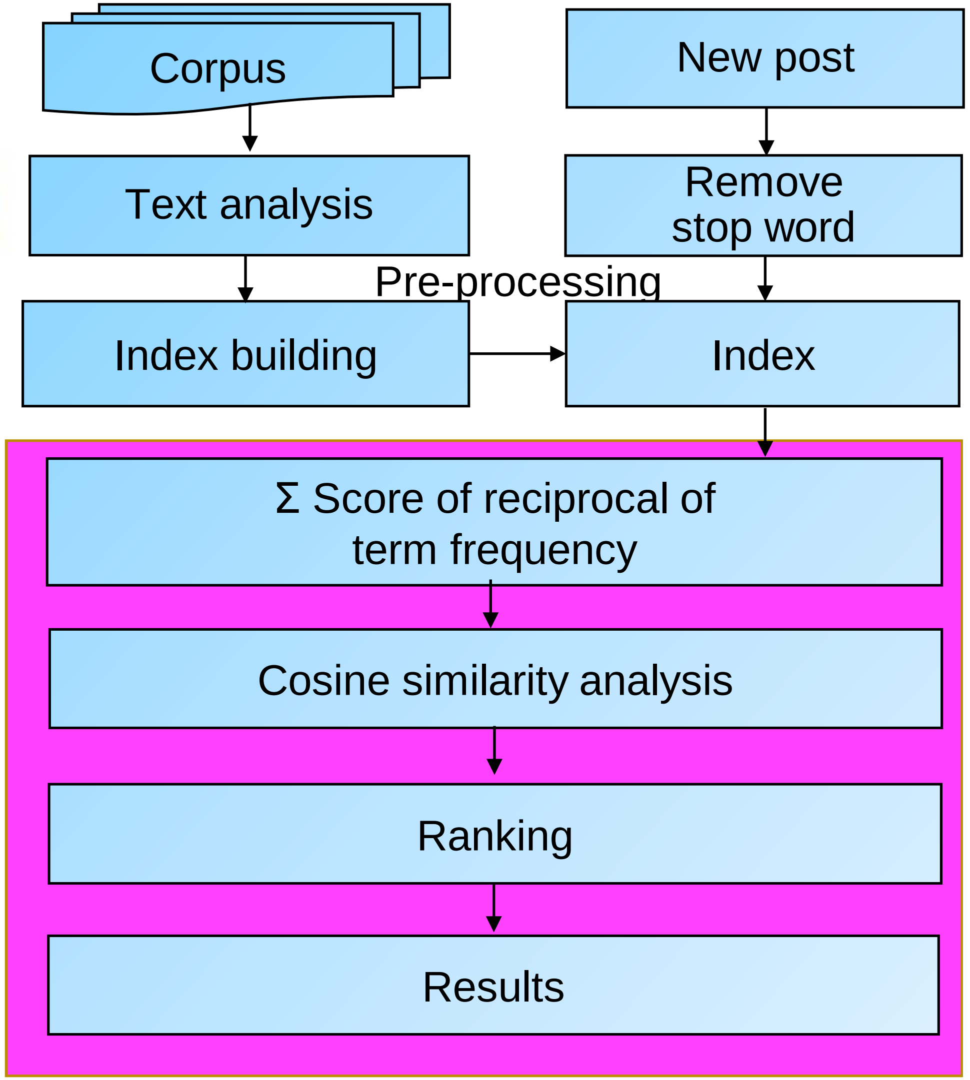


Retrieval-based Method: Search the new post



- When a new post provided, we searched the Solr index, and obtain the fetched potential candidate comments.
- We used all terms (words) from the provided new post one by one to search the Solr.
- If the term appeared in the post of post-comment pair, we fetched the "comment" (rather than post) as potential candidates for generated comments.
- Keep the first 500 search results

Ranking the Results



- We calculated the accumulated inverse term frequency.
- We computed the cosine similarity between the new post and the candidate comments.
- We multiplied accumulated inverse term frequency by cosine similarity as the relevance score.
- The candidate comment that match the assigned emotion and with highest relevance score was treated as the generated comment.

Evaluations | Retrieval-based Method

Evaluation Results								
Result	Submission	Method	Label 0	Label 1	Label 2	Total	Overall score	Average score
Evaluation result	RUN 1	Retrieval	716	200	84	1000	368	0.368

Only 3 teams submit for retrieval based method

Table 5. The result of the overall score and average score.

Team Name	Label 0	Label 1	Label 2	Total	Overall score	Average score
AINTPU_1	716	200	84	1000	368	0.368
IMTKU_1	580	248	172	1000	592	0.592
WUST_1	601	211	188	1000	587	0.587

Weakness of our retrieval method

We do not used semantic analysis before searching

- We used only the terms in the new post to search the results.
- We should also used similar term with similar meaning to search the corpus.

Emotion Categories

- We do not consider the noisy of emotion classification. We realize the precision issue of emotion categories after receiving the evaluation results.

Evaluations

| Retrieval-based Method

Evaluation Results									
Result	Submission	Method	Label 0	Label 1	Label 2	Total	Overall score	Average score
Evaluation result	RUN 1	Retrieval	716	200	84	1000	368	0.368

We realize the precision issue of emotion categories after receiving the evaluation results.

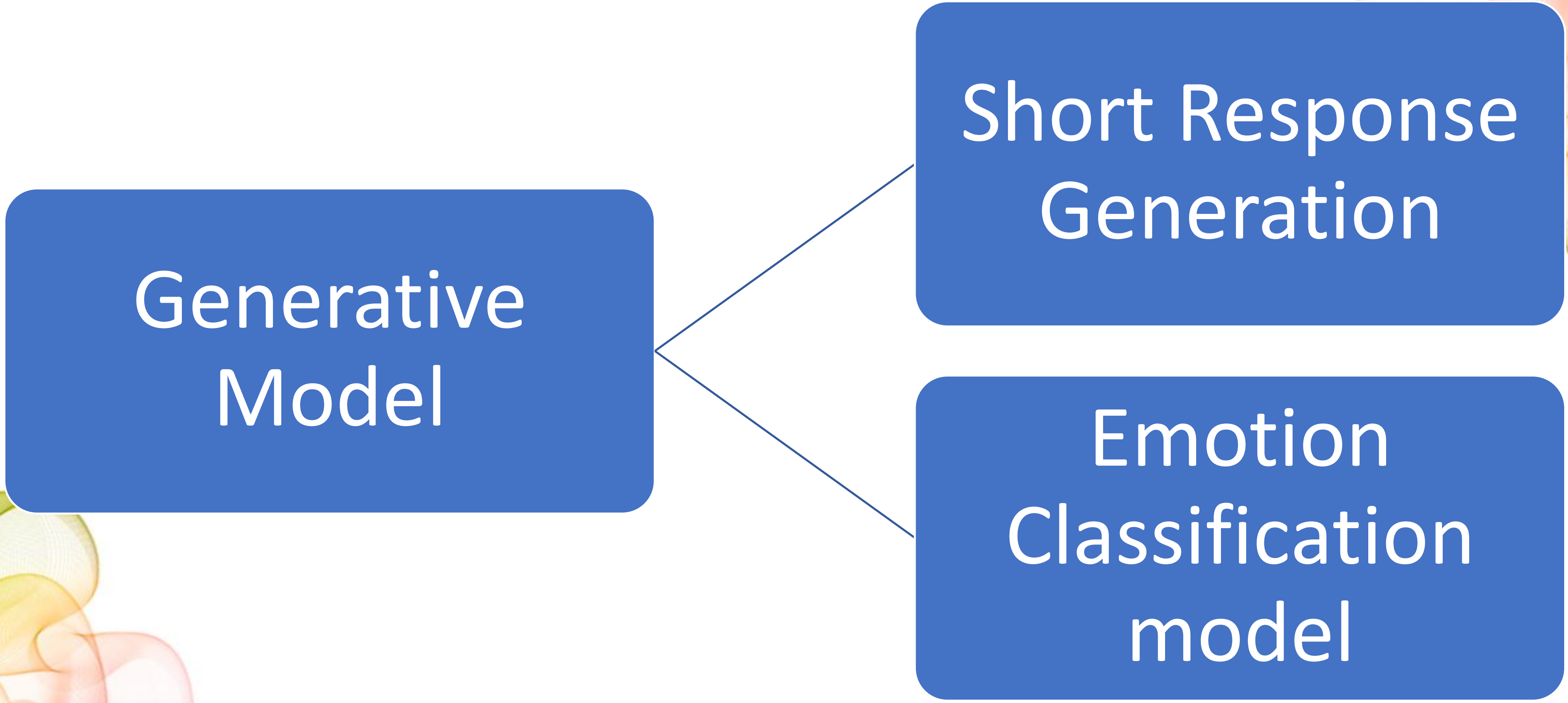
Only 30% (84/284) response were with correct emotion.

According to the organizers, the accuracy rate for emotion classification was 62% in their NLPCC papers. The actual accuracy rate may be lower than that.

Generative Approach

Automatically generate responses to questions

Generative Approach



Generative Models

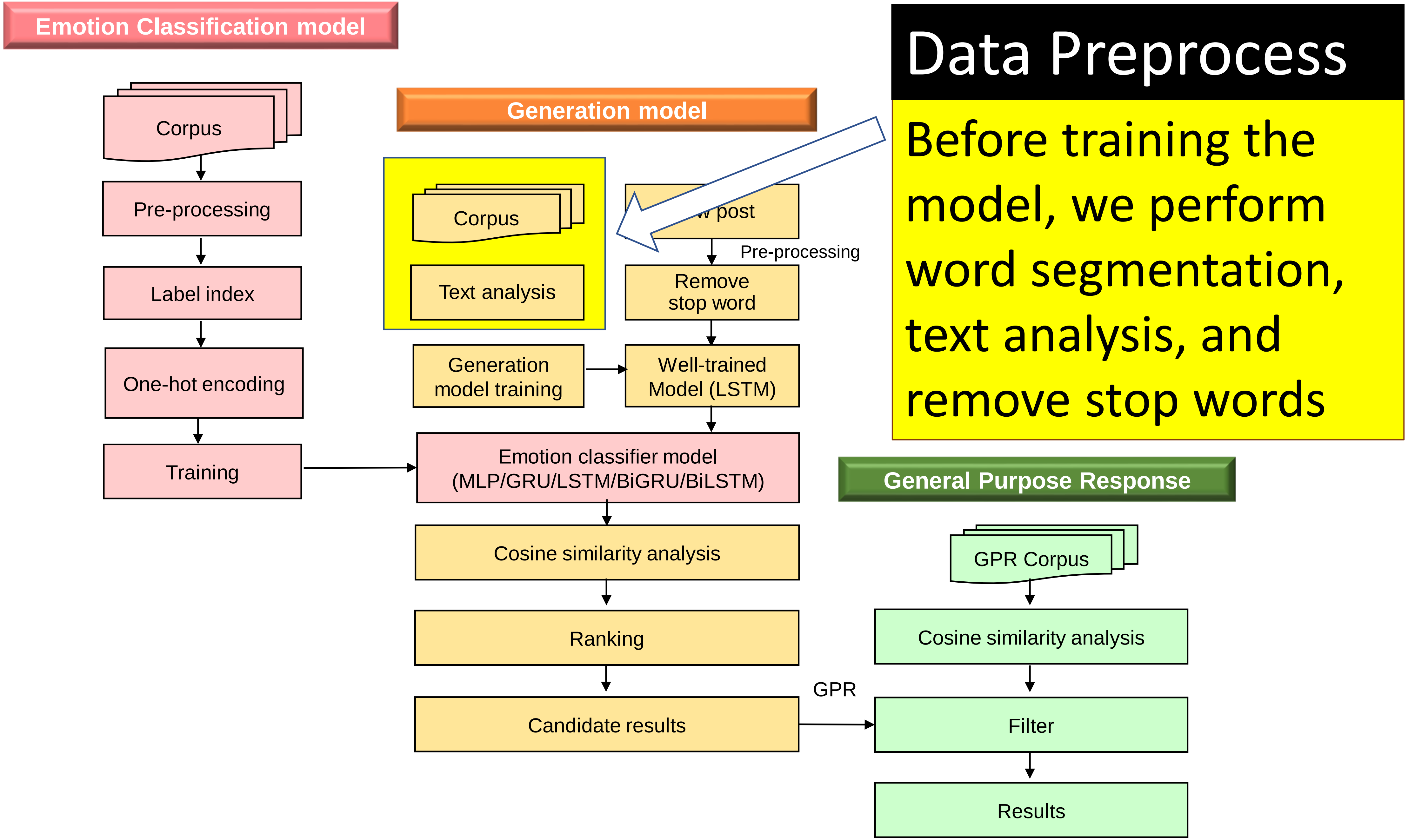
Automatically Generated Response in Short text conversion

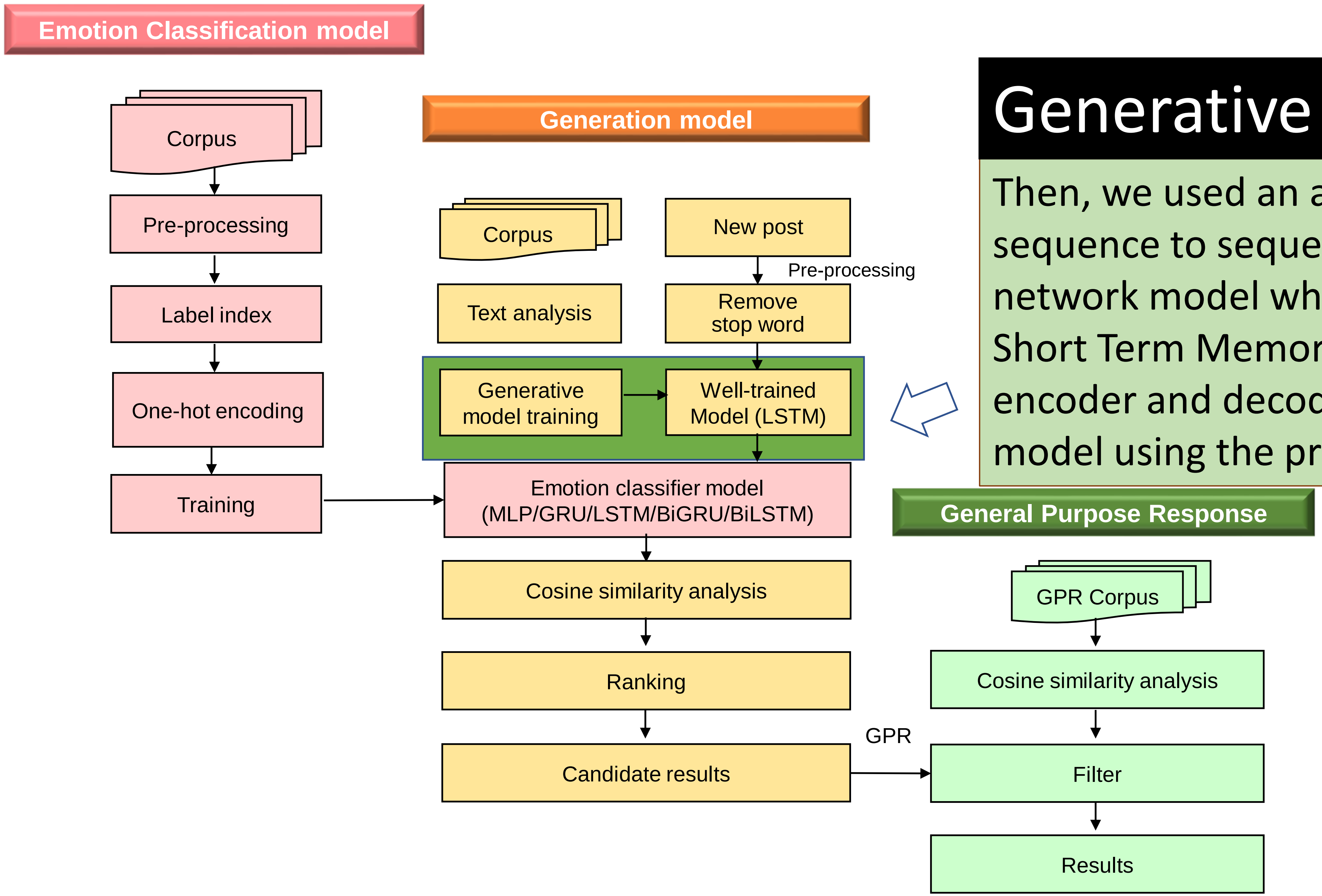
We employed an **attention-based sequence to sequence (Seq2Seq)** network model for the generation-based approach.

Generative Models | Generation-based Method

Generate Short Responses to the Dialogue

- Seq2Seq with attention mechanism
- Long Short Term Memory (LSTM) as encoder and decoder


Emotion Classification model
Corpus
Pre-processing
Label index
One-hot encoding
Training
Generation model
Corpus
Text analysis
w post
Pre-processing
Remove stop word
Generation model training
Well-trained Model (LSTM)
Emotion classifier model (MLP/GRU/LSTM/BiGRU/BiLSTM)
Cosine similarity analysis
Ranking
Candidate results
GPR
General Purpose Response
GPR Corpus
Cosine similarity analysis
Filter
Results
Data Preprocess
Before training the model, we perform word segmentation, text analysis, and remove stop words

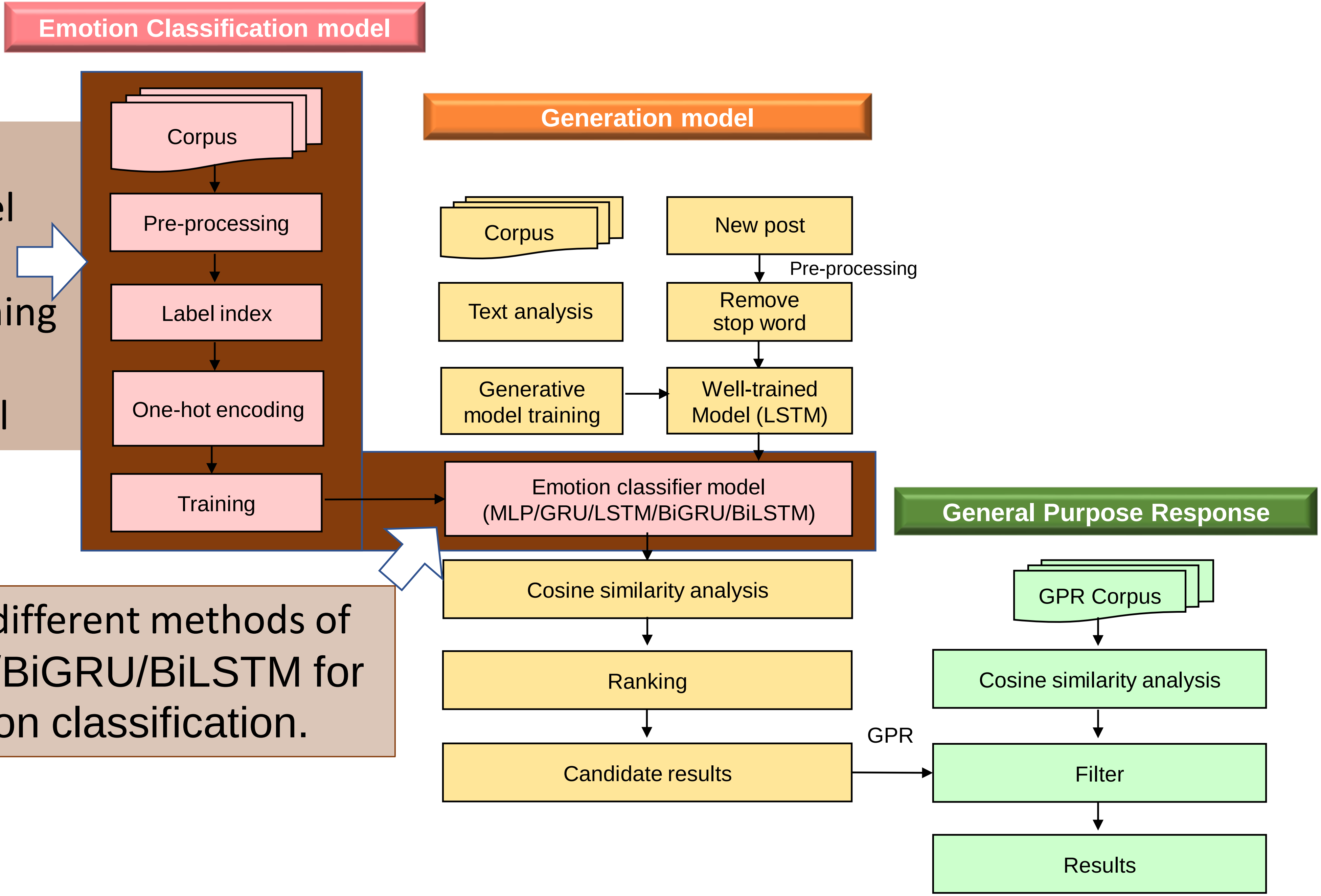


Generative Model

Then, we used an attention-based sequence to sequence (Seq2Seq) network model which take Long Short Term Memory (LSTM) as encoder and decoder to train the model using the provided corpus.

Emotion

We performed preprocessing, label indexing, one-hot encoding, and training to train emotion classification model



We compared the different methods of MLP/GRU/LSTM/BiGRU/BiLSTM for developing emotion classification.

Deep learning approach of Emotion Classification model

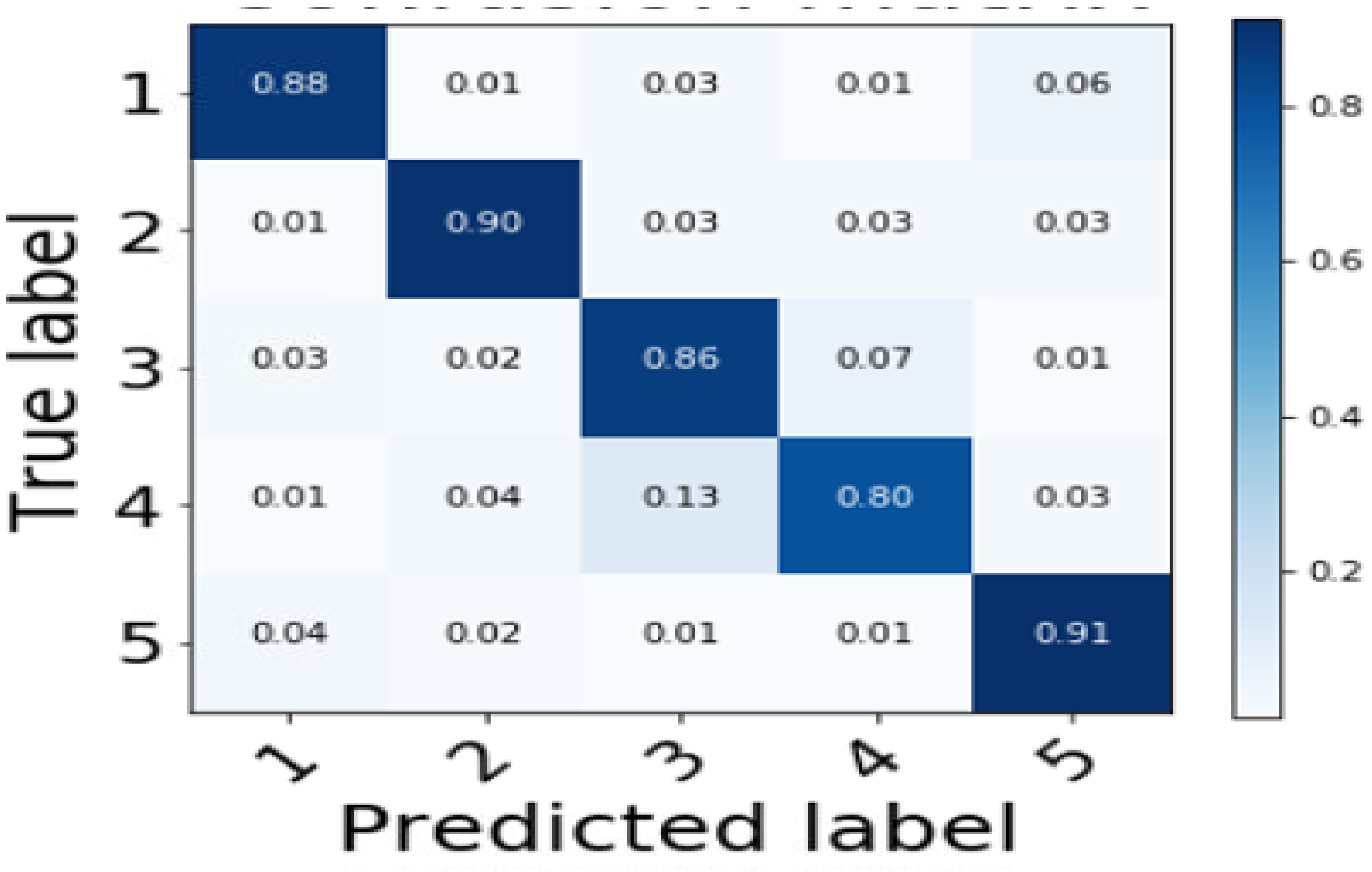
- MLP, GRU, LSTM, BiGRU, and BiLSTM

Evaluations of all all deep learning approaches

Evaluation Results					
DL model	Batch size	Dropout	Epochs	Accuracy	Loss
BiGRU	256	0.5	15	0.880	0.333
BiLSTM	256	0.4	10	0.879	0.335
LSTM	256	0.1	20	0.879	0.335
GRU	256	0.4	20	0.872	0.356
MLP	256	0.4	30	0.843	0.451

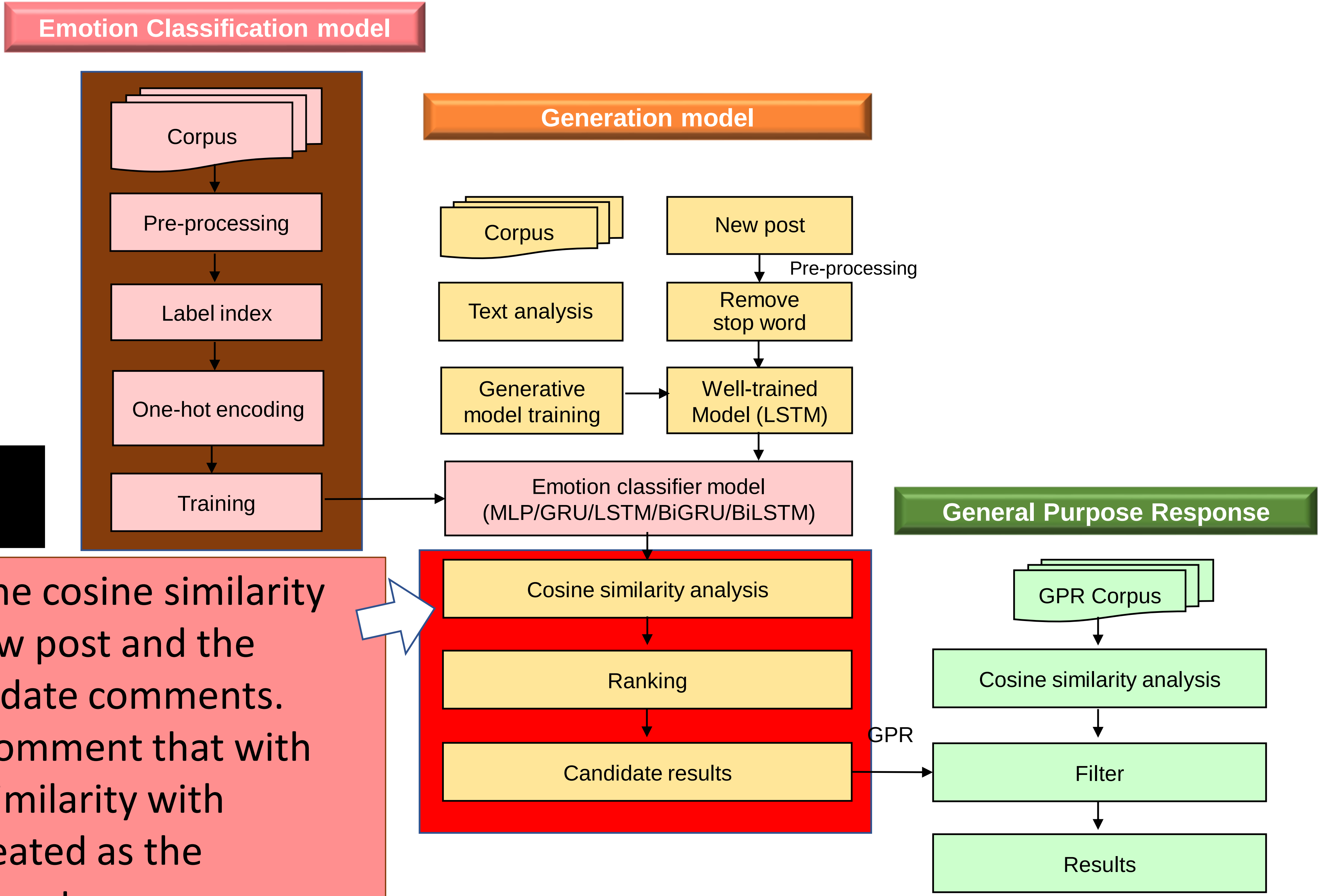
Confusion matrix for **emotion classification**

Best Method
Bi-GRU



Similarity

We computed the cosine similarity between the new post and the generated candidate comments. The candidate comment that with highest cosine similarity with question was treated as the generated comment.



Self-Evaluation Performance

Use MLP to automatically generate responses

Emotion classification	Label0	Label1	Label2	Total	Overall core	Average score
MLP	873	85	42	200	169	0.169
GRU	855	69	76	1000	221	0.221
BiGRU	860	72	68	1000	208	0.208
LSTM	864	65	71	1000	207	0.207
BiLSTM	857	84	59	1000	202	0.202

Self-Evaluation Performance

The emotion precision rate was only around 50%

Use MLP to automatically generate responses

Emotion classification	Label0	Label1	Label2	Total	Overall core	Average score
MLP	873	85	42	200	169	0.169
GRU	855	69	76	1000	221	0.221
BiGRU	860	72	68	1000	208	0.208
LSTM	864	65	71	1000	207	0.207
BiLSTM	857	84	59	1000	202	0.202

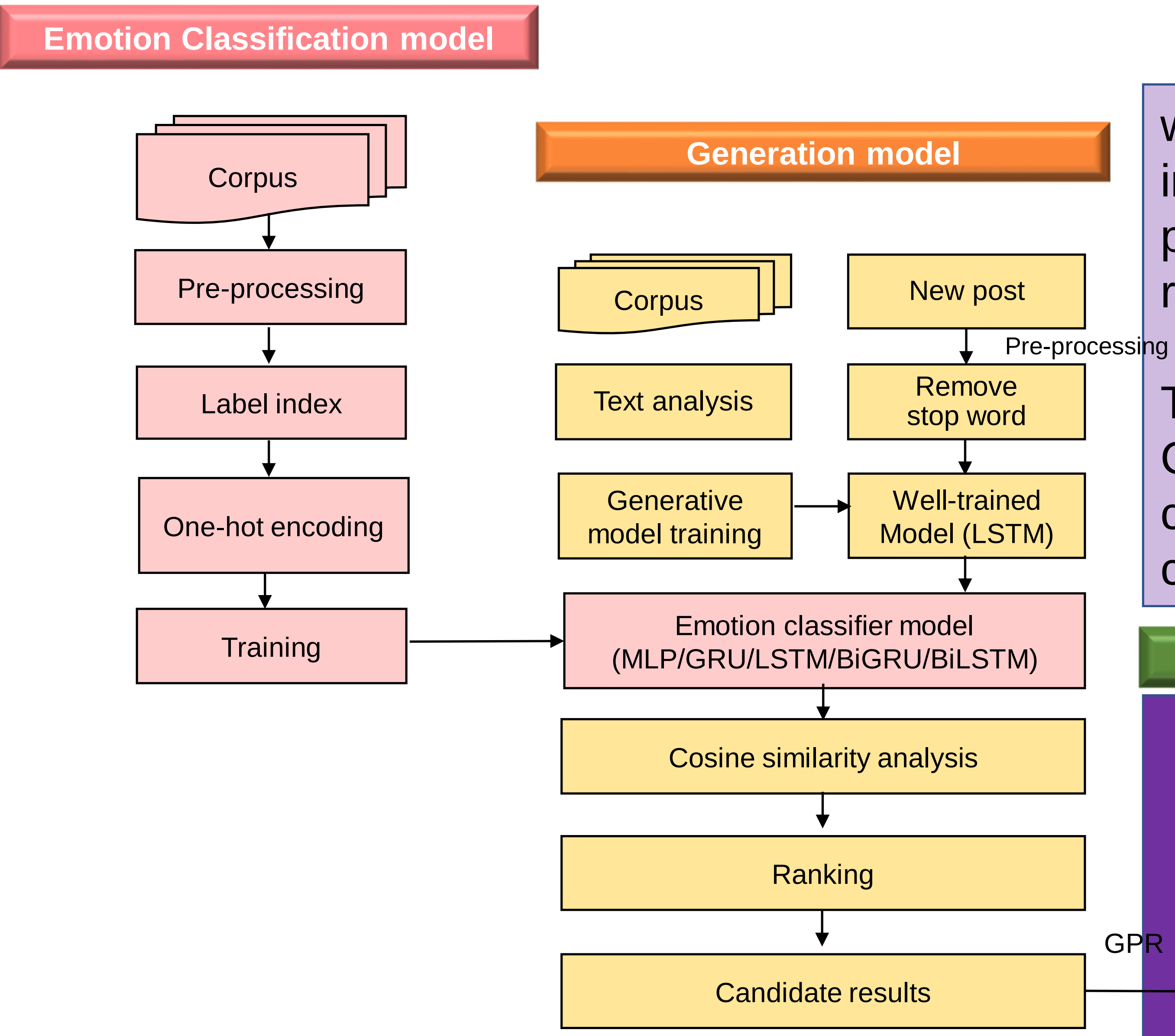
General Purpose Response

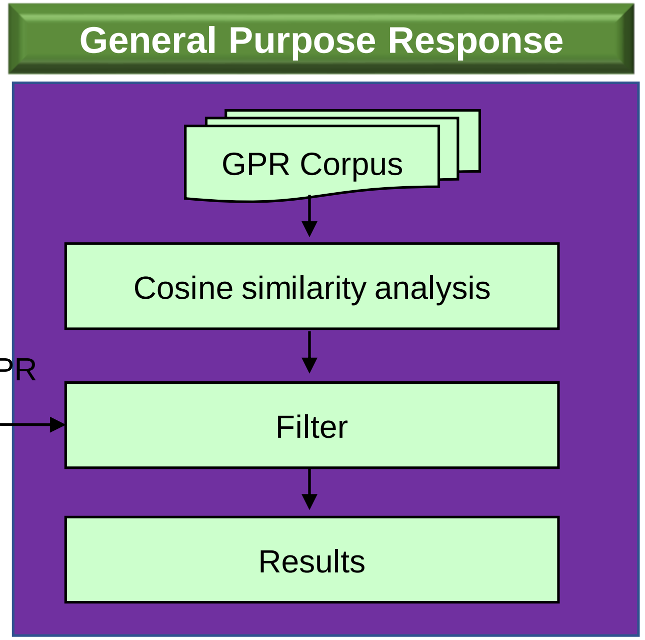
Generate responses when we do not know how to answer the questions

General Purpose Responses

we used **General Purpose Response(GPR)** to improve the generative-based response performance. About 1500 general purpose responses were created.

The generated comments will be replaced by the GPR at filter stage if the new post and generated comments received a low relevance score computed by cosine similarity (about 30%).





MLP+ General Purpose Responses

Use MLP plus GPR to automatically generate responses

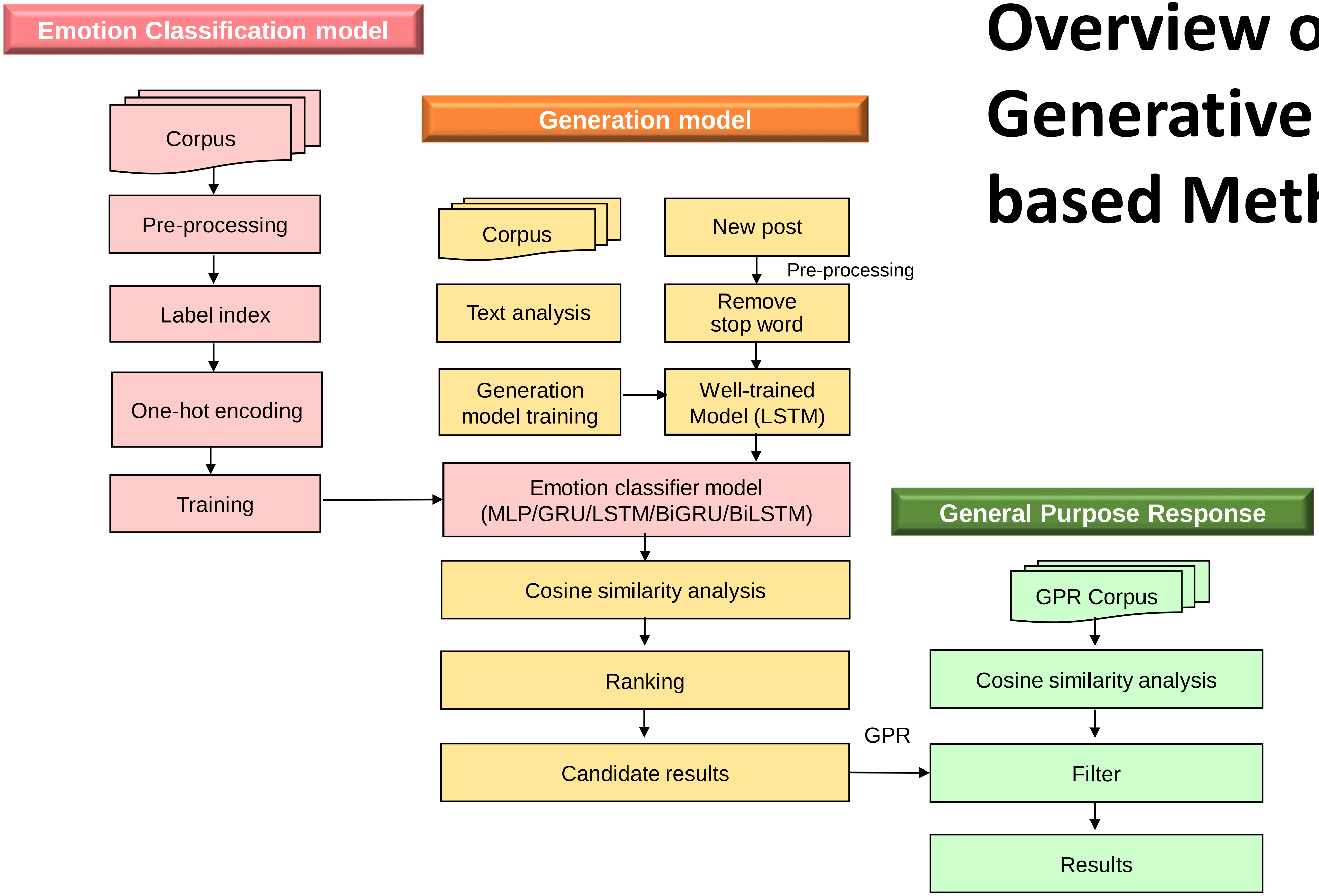
Emotion classification	Label0	Label1	Label2	Total	Overall core	Average score
MLP	808	124	68	1000	260	0.26
GRU	756	77	167	1000	411	0.411
BiGRU	727	111	162	1000	435	0.435
LSTM	749	89	162	1000	413	0.413
BiLSTM	753	75	172	1000	419	0.419

With or Without GPR

Use MLP to automatically generate responses

Emotion classification	With GPR Average score	Without GPR Average score	Difference
MLP	0.26	0.169	+0.091
GRU	0.411	0.221	+0.190
BiGRU	0.435	0.208	+0.227
LSTM	0.413	0.207	+0.216
BiLSTM	0.419	0.202	+0.217

Overview of Generative based Method



Conclusion

Comparison between methods

- Performance of Retrieval-based model is better than Generative model

- However, use different approach of deep learning in Emotion Classification model will have different kinds of improvement

- Further more, use EGPR can make performance more close to retrieval-based model

Evaluation of Emotion Classification model

- BiGRU > BiLSTM > LSTM > GRU > MLP

Future work

1. conversation model

- use seqGAN as deep learning neural network of generative model
- try to add topic layer between encoder and decoder of S2S architecture

2. EGPR

- take more general condition to expand EGPR dataset

3. Emotion Classification model

- Bidirectional Encoder Representation from Transformers (BERT) to improve the performance of emotion classification model