Report 3.4 on today's learnings: E1 Byeongjun Park

Today, we learned text mining. Before practicing code, we listened to the story explanation. I learned that there are language dependency, domain dependency, and amplitudes as challenges that can occur during text mining. And I learned the six processes of text mining in turn. When collecting text data, there were methods such as web crawling, and in this process, I learned what to be careful about. In the pre-processing stage, tokenization, lemmatization, stemming, and stop word removal were learned. And the learned theory could be visually confirmed through the code. Regarding vectorization, I understood the concepts of word embeds and word2vec that appeared after knowing that there was a problem of many 0 entries and a large vocabulary size during one hot encoding. It was amazing that words in similar vectors had similar meanings.

Next, we heard Daniel's explanation of text generation and image generation. It was a week-long class, but it was really nice to meet you. It was interesting to explain with image examples when explaining the diffusion model or image generation using dall-e. Finally, we had time to understand by using the transformer model and solving problems related to it.