Text mining of qualitative data in surveys

The background:

A market research company approached CoGo Data to assist them to automate the analysis of qualitative data.

The problem:

The company needed to generate insight from surveys conducted by their field workers. Most of the data consists of responses that are lengthy sentences that contain a lot of irrelevant or overhead information. Analysing the qualitative data is challenging because there is too much data to assign labels manually (thousands of responses for multiple questions). The fieldworkers conducted the surveys in their second language, resulting in spelling and grammatical errors, and some questions were open-ended (labels are not always apparent at first glance).

The solution:

The aim was to mine and assign a label to each response so that it can be aggregated to gain meaningful insights. CoGo Data developed a custom use-case data pipeline that utilises natural language processing to clean and tokenise the data, mine or identify the associated labels with a machine learning topic model approach, and assign and aggregate the labels to produce the relevant insights.

The result:

The data pipeline provided the company with the needed insights for its market research studies and similar questionnaires that they need to process in the future. The pipeline improved the company's productivity in terms of the analysis to a fraction of the time it would have taken to perform it manually.