

Project 4 – MVP

Domain

My goal is to categorize and cluster the different kinds of messages that different users would send, as well as request for, based on their profiles to find out the different users' interest in their messages. This information would be useful for chatbot customization – identifying the kinds of information users are interested in or whether they would use the bot over time.

Some of our features above and beyond from the text include gender, Facebook page, first contact date (of the bot), last contact date (of the bot), locale and date of message.

Data

The chatbot data is provided by prime.ai, which deploys various chatbots all over Facebook. It consists of about 2.5 million text messages sent by users to chatbots, as well as about 8 million other interactions with the chatbot. Below are two examples of messages sent to a chatbot:

Note that the bots created are very standardized and will only respond if users click a button, not by typing. The data does not contain the chatbot responses.

'I do not want to see jokes but I want to read any funny jokes.'
--

'Hi there are any other personal and business like me'
--

Known Unknowns

We don't know whether the user would successfully purchase the Facebook page's product or click on the link provided by the chatbot – i.e. data on whether the user clicked the links presented by the chatbot.