EPI-USE LABS
107 Boardwalk Boulevard
Boardwalk Office Park, Phase 5, Block B
Pretoria

value through innovation

Botic

14th March 2019

OVERVIEW

According to a Gartner report, by the year 2020, 85% of customers interactions will be managed without a human. Chatbots form a big part of customer interaction to assist a customer immediately, in a proactive and always-available manner. Sophisticated AI chatbots can be trained from historic data, to interact with a person in a specific context and thus providing value and insight to the customer.

Since the implementation of General Data Protection Regulation (GDPR) in the United Kingdom and shortly the Protection of Personal Information (POPI) Act in South Africa, it is crucial to NOT have any unnecessary personal information in your databases. It is a liability for your company and the risk of exposing the information could lead to serious financial and reputational damage.

Another area where personal information is often captured is when a company assists a customer through their online support system. These systems often called customer support ticketing systems, or simply ticketing system, capture interactions between the customer and a customer support representative in a forum like manner.

Considering both the idea of an AI chatbot being trained with historic data; in this instance with data from a ticketing system. In many instances, some personal information is captured in a ticket and then purely training a bot on that information could lead to the chatbot accidentally exposing some personal information about a customer. We have a need for a chatbot, that was trained on ticketing data, without the risk of exposing any personal information.

THE PROJECT

The project consists of two parts, both applied to our ticketing system. We would like to preemptively warn a customer if a message that they are going to post contains any personal information. The message should then be analysed and if the chatbot (trained on our historic ticketing data) can reply with a suggestion or follow up question, it should do so without exposing any personal customer information.

PROJECT SCOPE

1. The first part of the project is to ‘scrape’ a ticket post before a customer submits it, to warn if there are any private information the person is sending in the ticket.
2. The message from the customer should be interpreted, using machine learning (ML), in order to classify the content.
3. Using the aforementioned content classification, an intelligent bot service (trained on historic ticketing data) should then be able to respond to the ticket.
    a. The response should be contextual and possibly provide some suggestions to the customer,
    b. without accidentally disclosing any other private information that was part of the tickets that the bot was trained on.
    c. The bot response should also be ‘scraped’ to ensure no personal information slipped through. If an automated response is flagged as possibly containing personal information, the response should rather than an automated response, become a suggestion for a customer support representative.

ARCHITECTURAL REQUIREMENTS

- A simple demo app (we recommend Angular SPA)
- Application programming interface (API) service for the demo app to interact with the bot (we recommend Node.js)
- Use of natural language processing (NLP), we recommend word2vec built using Google’s TensorFlow, for processing the message for
    - containing personal information, and
    - assigning a content classification.
- Machine Learning (ML) or Deep Neural Network (DNN) for training the ticket bot
- Generating an automated response to a ticket post
    - based on the content classification, and
    - determining if the response contains and personal information in the message

SKILLS REQUIREMENT

Except for the skills obtained while studying Computer Science, no additional skills are required.

PROJECT DELIVERABLES

- Source code in a Git Repo

- Docker container with:
    - Demo app
    - Ticket bot service
        - to scan ticket messages for personal information,
        - to classify ticket content, and
        - to generate an automated response to the ticket message.
    - Ticket bot trainer
- Documentation (everything digital - save the 🌳)
    - Project plan
    - Requirement documents
    - Architectural design
    - User guide

EPI-USE LABS INTERACTION

The project owners can be contacted if the group has any urgent questions. Access to subject matter experts for each of the technologies can be arranged if the need arises. A timeline of the deliverables should be submitted with the project plan. Fortnightly virtual meetings should be scheduled with the project owners and monthly on-premise demo session to review progress made on the project.

PROJECT OWNERS

Jhani Coetzee
jhani@labs.epiuse.com

Tiaan Scheepers
tiaan@labs.epiuse.com

FYI

- https://github.com/facebookresearch/fastText
- https://www.youtube.com/watch?v=2V6NHKmfnW0