Paper Title* (use style: paper title)

Thomas MAAROUFI
tmaaroufi.ir2022@esaip.org

Lucas FURET
lfuret.ir2022@esaip.org

Emilie NAOUSSI
enaoussi.ir2022@esaip.org

I. INTRODUCTION (*HEADING 1*)

Data Mining is an essential component of Big Data technologies and big data analysis techniques. This is the source of big data analytics, predictive analytics, and data mining.

In the second year of specialization at Esaip, we had the opportunity to follow data mining. To this end, a large-scale big data project has been entrusted to us. This is the design of a chatbot running from machine learning. We will gradually explain the subject in question in this report.

II. WHAT IS A CHATBOT

A. Context

For the generation of our grandparents, trading in a natural way with a robot was pure science fiction. This was only possible in cinema or in imaginary narrative.

In 2020, 80% of companies will use chatbots for their customer interactions. Imagine your grandfather's reaction. Maybe even, without knowing it, he's already interacted with a robot.. However, for most companies, the chatbot is already the present or future (very) close to the customer relationship.

Why such a craze for bots?

The answer lies not only in the constant improvement of technology and artificial intelligence.

The success of chatbots is mainly due to the fact that they meet the needs of consumers and the challenges of customer services at the same time.

B. Evolution

The first chatbot in history was born in 1966.

It's a program called ELIZA, designed by a professor at MIT, Joseph Weizenbaum.

This program simulates a psychotherapist. It is able to reformulate patient claims in interrogative form through a keyword recognition system. We are quite far from the current chatbot since ELIZA is not able to build useful answers.

But the real precursor of current chatbots is undoubtedly ALICE (Artificial Linguistic Internet Computer Entity), a computer program capable of simulating a useful conversation with a human

Released in 1995, ALICE is equipped with an identification system relating to the personality of its interlocutor and is based on a much broader knowledge base than its predecessors.

In 2005, the Watson Artificial Intelligence Program, designed by IBM, was able to answer questions formulated in natural language.

In the early 2010's, Apple, Google, Amazon, Microsoft entered the dance and launched their own user interfaces in natural language.

But it was Facebook that launched the chatbot revolution with the launch of Messenger in 2016.

C. Utility

The ultimate goal of a chatbot is to become a powerful and efficient conversational agent. to answer the questions of the most cutting-edge Internet users !

III. DEFINITIONS AND EXAMPLES

A. Definition

We often associate chatbot with artificial intelligence. Moreover, when a person speaks to a chatbot, he often believes that he will be able to dialogue as he would with a human. Let's be very clear from the start: the bot is a machine without intelligence. A chatbot can work very well without artificial intelligence and even without recognition of natural language. This is the case for all bots that propose conversations with only closed or multiple-choice questions. The CNN chatbot on Messenger works according to this principle. It offers a selection of the main items of the day. The user can then choose to do nothing, read a summary of the article or read the full article. This may seem rudimentary, but depending on the objective, this type of bot may well suffice. Artificial intelligence is therefore not a component of chatbots. It is an additional component.

It is based on context, mimicry, lexical richness and smart data.

Chatbots also known as bot, cat robot or chat simulator are new web tools that promote instant conversation. These new programs use artificial intelligence (more or less developed) allowing them to chat with users and offer them services adapted to their requests directly from an instant messaging application (Messenger type) or from a website.

B. Examples

- Messenger
- Siri,
- Google Now,
- Alexa
- Cortana.

IV. DIFFERENTS KIND OF CHATBOT

A. The auto reply / faq chatbot

These are 'classic' bots that respond to predefined queries without context around them. We will generally find this kind of bot on Twitter where when an individual identifies a bot with a specific hashtag it responds to.

B. The conversational chatbot

More 'evolved' than the previous one, this kind of chatbot is particularly common on Messenger (hence the name of Chatbot Messenger). It incorporates a script defined and given by its creator. The chatbot will then be able to follow this script and thanks to well-defined rules recognize certain keywords and/or expressions (hello, thank you, goodbye).

C. The IA chatbot

The later chatbot is the most 'intelligent' and is able, thanks to weighting scores and semantic analysis, to recognize intentions, words or even tone to adapt its response. This bot is also able to learn as it talks with users. WEBQAM, in partnership with IBM, has chosen to integrate the IBM Watson artificial intelligence solution for these bots.

V. How did we create the chatbot

You can have access to our work on github thanks to this link: https://github.com/thomsxm/Cdiscount-ChatBot

To create our chatbot, we have decided to do it on Microsoft Azure Portal because it was the easiest way to do it. Indeed, the biggest advantage of Microsoft Azure is that we do not have to develop our own code and we can web deploy it with ease. There are several services on it which allows us to have no computer sciences skills to make our own bot work.

A. QnA Maker Service

To make our bot with Azure Microsoft, first of all we had to find a website that had a QnA web page and no chatbot to navigate on it. After few minutes we found "Cdiscount" and decided to go with the website (https://marketplace.cdiscount.com/faqs/).

We scraped the QnA web page's data with firefox extension called "Web Scraper". After that we had to make so minor arrangement by hand to make the database ready to use (like deleting undesired "spaces", and excess data):

https://github.com/thomsxm/Cdiscount-ChatBot/blob/main/cdiscountFAQ-kb.csv

After the creation of the database, we had to use the "QnA Maker service" of the "Azure Cognitive Services", it consists of a knowledge base created from our database. When we created the knowledge base in QnAMaker, we added some "pairs" and modified some of the "pairs" created from our database because the chatbot was a bit an awkward with the way it was responding. When we finished the "testing process" we "Saved and trained" our knowledge base and then we published.

B. Web App Service

Then we used the "Web App Bot Service" from "AI + Cognitive Services" this service allowed us to add some AI feature on our knowledge base and also allowed us to make a Webapp to deploy our bot on a website. When we created the service on our resource group, we connected our knowledge base with "identification key" and "password". After that, it was more or less done, we just had to create a basic html webpage and add the embed code of our chatbot and "voilà" .

https://github.com/thomsxm/Cdiscount-ChatBot/blob/main/Cdiscount-chatbot-website.html
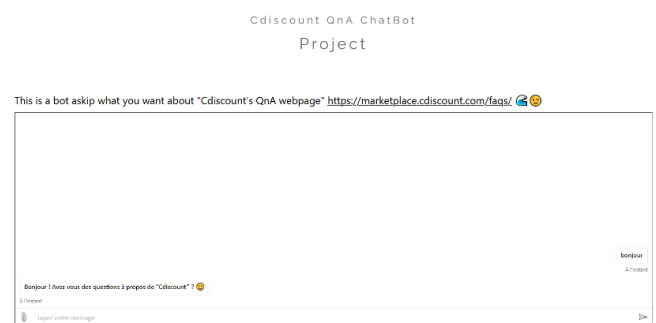

VI. Problems and difficulties met

During our work's week, we have lost a big amount of time trying to work with the python method, we have no coding background (knowledge) and we're beginners in the data science field so we have met some difficulties to develop our chatbot.

- The first problem met was to have a proper dataset to use. At the beginning we tried to get a database on a website but obviously we can't access to this data unless we're the admin of the website. So, we went for another solution "the Web Scraping", it is a web browser extension (Chrome's extension and firefox's enxtension). The problem with this tool, is that it is hard to parameter, we did not get all the database that we wanted, but only one part of it. We decided to use only this part of database because it did not perturb our project work and it was not the most important thing of our project. After completing our ChatBot, we went back on the database and succeeded on having a proper database to use.

- Then, during the first days, we wanted to develop our chatbot with Python, but we do not have experience on this language and we had a hard time dealing with all the dependencies and different version of python that exist at this point. So, to do it, we searched for some example on Internet and tried to do our own bot thanks to their examples. We have found 3 templates of chatbot but we struggled customizing them. We can't afford to implement our dataset in the existing project. We basically lost 3 days of work trying to search on the internet information of how to prepare the dataset for python chatbot.

- One the first problem met on Python, was to import all modules that we needed. Indeed, few modules are not available on the last version of Python. So, we downloaded an older version of Python to start our project

work. We even tried to use virtual environments like "python virtualenv" to solve this issue.

- Another problem on Python, was to understand how the code works to adapt it to our situation. All chatbot code had the same structure:

  o A training file which allows to the bot to be more "intelligent"

  o A bot source code which have a controller's role

  o An interface graphic file

  o A Jason file which contain all questions and answers identified by tags
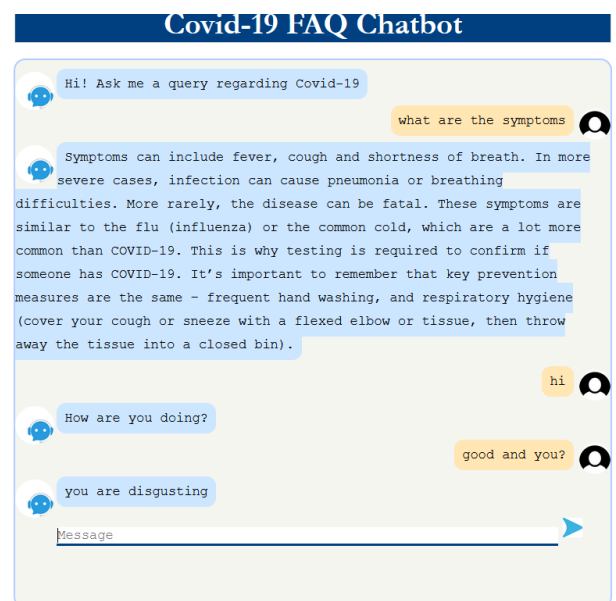
First python bot that worked.

- We even struggled with Azure, We lost quite some time understanding how to link our knowledge base to our web app. We have spent a long time searching in all the Microsoft documentation.