

vtn180004

CS 4395.002

### Vincent Van Gogh Chatbot Report

For the chatbot project, I decided not to use any frameworks such as Amazon Lex or Google's DialogFlow, since those frameworks would do the majority of the coding. So, I decided to design and implement an interactive chatbot from scratch. The chatbot I created is specifically built to communicate with users about information on Vincent Van Gogh, one of the most famous Dutch painters. The user will be able to ask questions or make direct statements to the prompt and the chatbot will formulate a response based on the knowledge base. If it's not in the knowledge base, the chatbot will respond accordingly related to the intents of the user. The chatbot will keep track of different users' data/information by creating a user model for each user. The user model will store each user's name as well as remembering the different topics that the user is interested in or not. The information of the model will then be stored into a pickle file.

First, the system initializes by opening the intents JSON file, which is how most of the responses that are outside of the knowledge base will be determined between the user and the bot. Within the intents file are collections of 'intents' with different categories such as greetings, welcome\_back, goodbye, thanks, likes, dislikes, etc. The bot will determine which intent is relevant based on a list of user responses. If there is a certain threshold of similarity between the user's response and the list of possible patterns, that are within the intent, is reached. Then, the chatbot will use the specific intent's list of responses to message back to the user. The classify function performs this calculation, using a model with the x and y values representing the words and different intents respectively. Following the entry of those data into Sklearn, the model will assign the most relevant intents and is sorted by the strength of the probability. Another thing to

mention is that within the intents JSON file, the "tags" has a list of context\_req, which describes what prerequisites must be satisfied in order to access this intent, and context\_set, which will set the context once the intent is reached and used.

After the JSON file is loaded, then the neural networks model that was created using sklearn, will be loaded in. To improve the performance of the model, I applied several natural language processing techniques to the training data. All the stop words were removed, with the exception of no, not, can, are, do, okay, like, know, am, who, want. The next step was to lowercase and stem each word. After eliminating all duplicates, I had a better collection of data to work with and the user data, which is represented as a pickle file that is constantly being updated, is then loaded and initialized. This will enable the chatbot to ask for the user's name to start the conversation. The chatbot will select the most relevant intent when the user starts communicating with it and will base its response on that intent. The intent system is exclusively used for the communication between the user and chatbot when given simple questions that are beyond the scope of the knowledge base. When the system detects that a user is starting to ask questions about the topics of Van Gogh, the system would flag that it will need to go into the knowledge base to give a response. In this case, the user's requested subject is saved and subsequently lemmatized. After that, the system would use the lemmatized word to look for the topic in the knowledge base. Then, extract the sentence that is relevant from the question/request and return it to the user.

During the process, the system will also record and categorize the user's statements as either positive, negative, or objective. When a user makes a positive statement, it would get added to the user model as a topic of interest. If it's a negative statement, the system would remove that from the user model if it had been a topic of interest. If it is an objective statement,

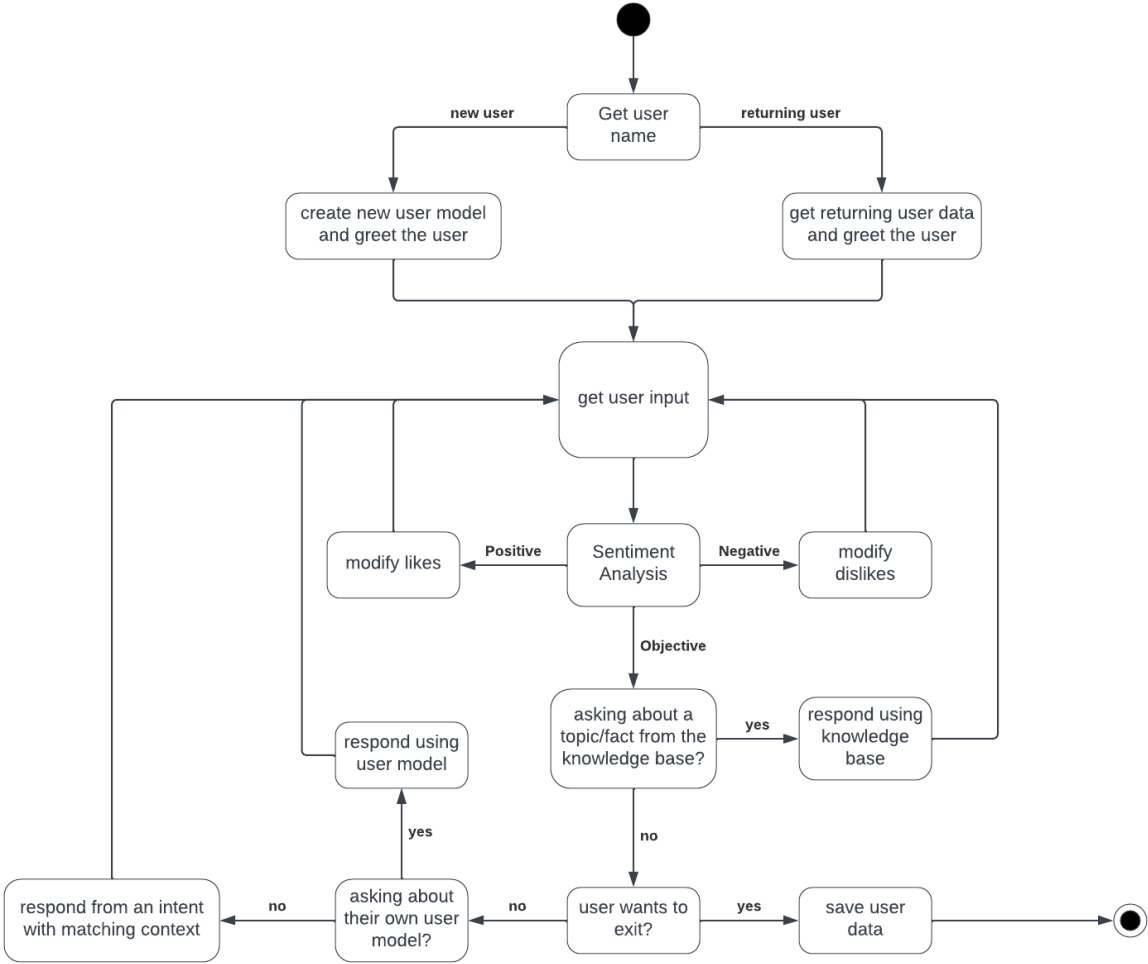
the system will proceed without changing anything. This is determined by applying sentiment analysis, one of the natural language techniques that we have learned. The method I chose for doing this is called VADER (Valence Aware Dictionary and Sentiment Reasoner).

I used the web crawler that I had developed from the previous project to build the knowledge base, and rewrote some of its code to make it specifically for the topics of Van Gogh. The web crawler pulls relevant information on Van Gogh from different web pages, although, I manually changed some of the text in the corpus. Then, using the NLP technique, TF-IDF (Term Frequency - Inverse Document), the top 50 words from the web pages were selected. Afterwards, I personally selected the top 10 words from the list that I believed were the most relevant and interesting to use for the Van Gogh Bot. The words that the chatbot will focus on when conversing with it are: sunflowers, asylum, museum, nature, painting(s), night, brother, artist, works, and Gogh. Each of the key topics will be associated with a list of relevant responses on the subject, and the knowledge base is stored in a pickled dictionary. See the appendix on the knowledge base for a sample of one of the topics.

### **Evaluation of the Chatbot**

The chatbot was created from scratch without using any preexisting framework, which leaves the system quite open to errors. There might be a problem with the responses from the chatbot because the text that came from sites using the web crawler had weird symbols within the sentences. For instance, the bot may respond with, “the painting is believed to have been created by van gogh in 1888 â€” around the same time that his artwork "sunflowers" was made â€” just two years before his death.” Overall, the chatbot was able to carry on a limited conversation using the knowledge base and can save the users likes and dislikes.

# Diagram of Dialog Tree/Logic



## Sample Dialog Interactions

Chatbot

Van Gogh Bot

Hello! What is your name?

>>vicky

I'm glad you're back!

>>can you teach me

I know a few topics about Van Gogh: sunflowers, asylum, museum, nature, paintings, night, brother, artist, works, gogh

>>can you tell me more about sunflowers

van gogh painted two series of sunflowers in arles, france: four between august and september 1888 and one in january 1889; the versions and replicas are debated among art historians.

>>what am i interested in

I believe you are interested in starry night

>>what do i not like

I believe you aren't interested in asylum

Send

Hello! What is your name?

>>Momo

Hello ~

>>what can you do

I know a few topics about Van Gogh: sunflowers, asylum, museum, nature, paintings, night, brother, artist, works, gogh

>>i love sunflowers

I'm happy to hear that! I'll keep that in mind!

>>although, i hate asylums

I'm sorry to hear that. I'll keep that in mind!

>>what do i like

I believe you are interested in sunflowers

>>what do i hate

I believe you aren't interested in asylums

>>laskdhfkl

Sorry, I don't understand. Do you want to learn something about Van Gogh?

>>ye

I know a few topics about Van Gogh: sunflowers, asylum, museum, nature, paintings, night, brother, artist, works, gogh

>>tell me something related to his paintings

before coming under the possession of the van gogh museum, a norwegian industrialist owned the painting and stored it away in his attic, having thought that it wasn't authentic.

```
>>idk
Didn't quite get that. I can tell you some stuff about Van Gogh, would you like th
at?
>>ye
I know a few topics about Van Gogh: sunflowers, asylum, museum, nature, pai
ntings, night, brother, artist, works, gogh
>>tell me something about his brother
though he had no formal art training, his brother theo offered to support van go
gh financially.
```

## Appendix for Knowledge Base

```
NLP_Portfolio > Chatbot > dict_output.py > ...
1  # print to a file an appendix of the knowledge base
2
3  import pickle
4
5  dict = pickle.load(open('knowledge_base.p', 'rb'))
6
7  with open("appx_knowledgebase.txt", "a") as f:
8      for key in dict.keys():
9          print('Facts About: ', key, file = f)
10         print(dict[key], file = f)
11         print('-----', file = f)
12
13
14
```

```
NLP_Portfolio > Chatbot > appx_knowledgebase.txt
1  Facts About: sunflowers
2  ['vincent van gogh was one of the world's greatest artists, with paintings such as 'starry night' and 'sunflowers,' though he was unknown until after his death.', 'van
   gogh painted two series of sunflowers in arles, france: four between august and september 1888 and one in january 1889; the versions and replicas are debated among art
   historians.', 'the oil paintings on canvas, which depict wilting yellow sunflowers in a vase, are now displayed at museums in london, amsterdam, tokyo, munich and
   philadelphia.', 'it features self-portraits, "the potato eaters," "the bedroom" and "sunflowers.", 'the painting is believed to have been created by van gogh in 1888
   - around the same time that his artwork "sunflowers" was made - just two years before his death.', 'these were exceptional years for the master, showcasing his talent,
   as well as his torments, in iconic works such as sunflowers, irises, wheatfield with crows, the starry night, in addition to his japanese influences with the
   magnificent almond blossoms or his later the church at auvers.', 'the pictures he created over the following 12 months-depicting blossoming fruit trees, views of the
   town and surroundings, self-portraits, portraits of roulin the postman and other friends, interiors and exteriors of the house, sunflowers, and landscapes-marked his
   first great period.', 'as he was preparing a room in his yellow house to accommodate gauguin, who planned to visit arles for an extended stay later in the year,
   vincent van gogh was also working on his second series of the sunflowers paintings.', 'in 1888 van gogh settled in arles in provence, where he was visited by gauguin
   and painted his now famous series of 'sunflowers'.']
3  -----
4  Facts About: asylum
5  ['van gogh painted "the starry night" in the asylum where he was staying in saint-remy, france, in 1889, the year before his death.', 'in 1889, after entering an
   asylum in saint-remy, france, van gogh began painting irises, working from the plants and flowers he found in the asylum's garden.', 'van gogh decided to move to the
   saint-paul-de-mausole asylum in saint-remy-de-provence after the people of arles signed a petition saying that he was dangerous.', 'theo, who was suffering from
   syphilis and weakened by his brother's death, died six months after his brother in a dutch asylum.', 'a few months later, he voluntarily entered the asylum of
   saint-remy-de-provence, where he painted fiercely.', 'van gogh stayed there for 12 months, haunted by recurrent attacks, alternating between moods of calm and despair,
   and working intermittently: the starry night, garden of the asylum, cypresses, olive trees, les alpilles, portraits of doctors, and interpretations of paintings by
   rebrandt, delacroix, and millet date from this period.', 'subsequently, the artist entered care at the asylum at saint-remy-de-provence, and although struggling with
   rapid mental deterioration, he produced a series of stylistically diverse paintings that depicted the hospital itself and well as the surrounding grounds, olive trees
   and cypresses.']
6  -----
7  Facts About: museum
8  ['the painting is currently housed at the museum of modern art in new york, ny.', 'the oil paintings on canvas, which depict wilting yellow sunflowers in a vase, are
   now displayed at museums in london, amsterdam, tokyo, munich and philadelphia.', 'van gogh's self-portraits are now displayed in museums around the world, including in
   washington, d.c., paris, new york and amsterdam.', 'in 1973, the van gogh museum opened its doors in amsterdam to make the works of vincent van gogh accessible to the
   public.', 'the museum houses more than 200 van gogh paintings, 500 drawings and 750 written documents including letters to vincent's brother theo.', 'in september
   2013, the museum discovered and unveiled a van gogh painting of a landscape entitled "sunset at montmajour.", "before coming under the possession of the van gogh
   museum, a norwegian industrialist owned the painting and stored it away in his attic, having thought that it wasn't authentic.", 'at the museum i thought a lot about
   delacroix.', 'source status: original manuscript location: amsterdam, van gogh museum, inv.', 'they are now in the van gogh museum.', 'there were two paintings by
   adriaen brouwer, seven by adriaen van ostade, two by isaac van ostade and three by gerard ter borch in the rijksmuseum.', 'the sketch by peter paul rubens is christ
```

## Appendix for User Models

```
NLP_Portfolio > Chatbot > users > ⚙️ usr_samples.py > ...
1 # user model samples to test the chatbot
2
3 import pickle
4
5 class User:
6     def __init__(self, name, likes, dislikes):
7         self.name = name
8         self.likes = likes
9         self.dislikes = dislikes
10
11 # user samples
12 usr1 = User('Vicky', ['starry night', 'sunflowers'], ['asylum'])
13 usr2 = User('Shu', ['paintings', 'asylum'], ['sunflowers'])
14
15 users_data = [usr1, usr2]
16
17 pickle.dump(users_data, open("users_data.pkl", "wb"))
18
```

```
NLP_Portfolio > Chatbot > users > ⚙️ save_usr_info.py > ...
1 # save user model info to a file
2
3 import pickle
4
5 class User:
6     def __init__(self, name):
7         self.name = name
8         self.likes = []
9         self.dislikes = []
10
11 usr_info = pickle.load(open("users_data.pkl", "rb"))
12
13 for usr in usr_info:
14     print('User Name: ', usr.name)
15     print('Likes: ', usr.likes)
16     print('Dislikes: ', usr.dislikes)
17     print('-----')
18
```

```
$ py save_usr_info.py
User Name: Vicky
Likes: ['starry night', 'sunflowers']
Dislikes: ['asylum']
-----
User Name: Shu
Likes: ['paintings', 'asylum']
Dislikes: ['sunflowers']
-----
User Name: Erin
Likes: ['sunflowers']
Dislikes: ['asylums']
-----
User Name: Yoko
Likes: ['paintings', 'gogh']
Dislikes: ['brother']
-----
```