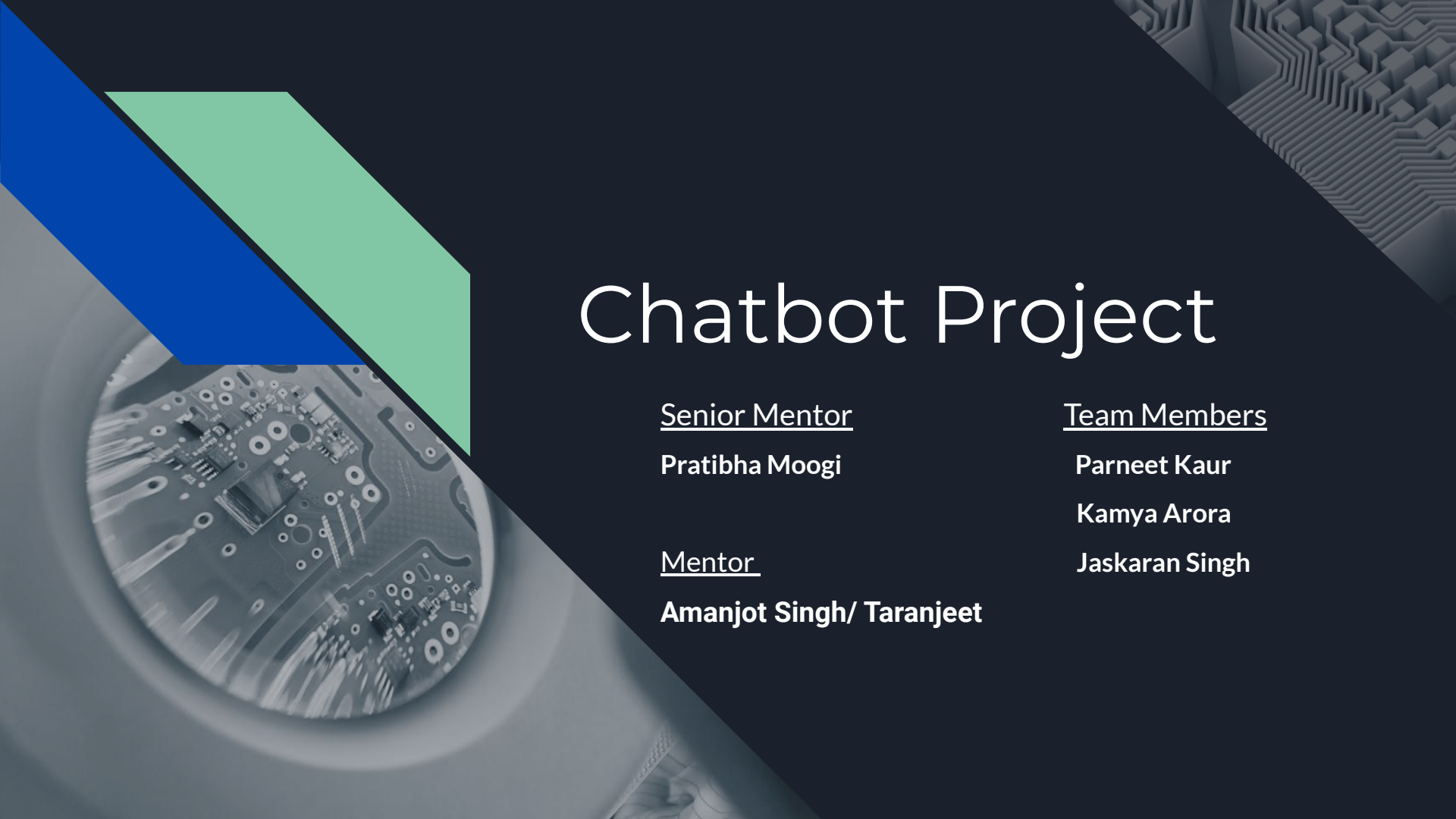


Chatbot Project

Senior Mentor

Pratibha Moogi

Mentor

Amanjot Singh/ Taranjeet

Team Members

Parneet Kaur

Kamya Arora

Jaskaran Singh

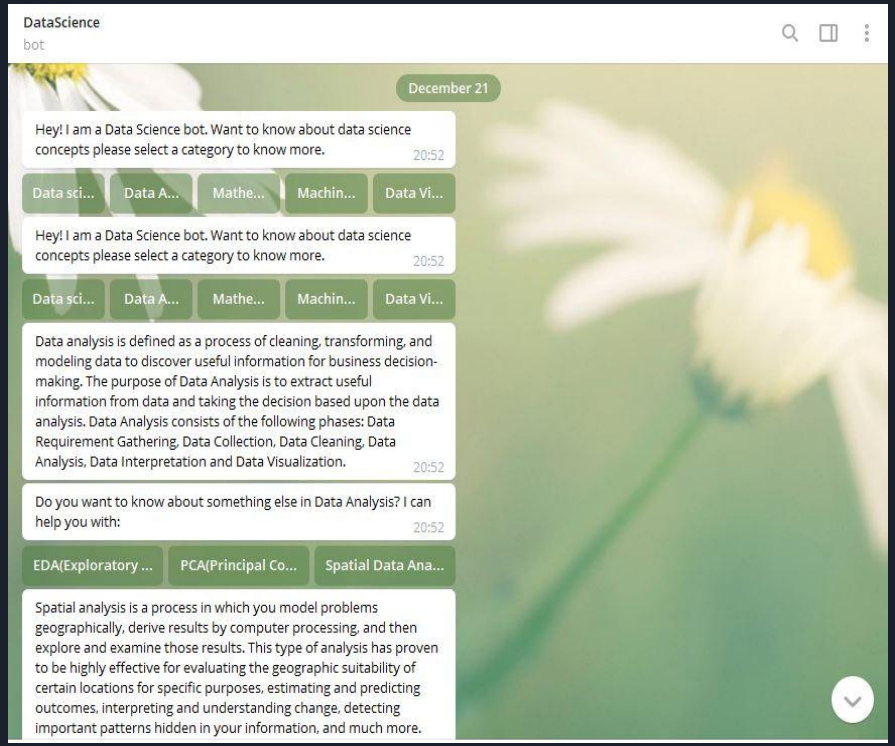
Introduction to Project

Chatbot is a computer application used to conduct chat conversations with machines in natural language. It is widely used in MNCs that earlier had a large number of human workforce to handle customer support. It is the most advanced and popular application of Human Machine interaction.

Our goal is to make a Data Science Bot that can answer the questions related to the field of Data Science as per the requirements of the user.

Objectives of the Project

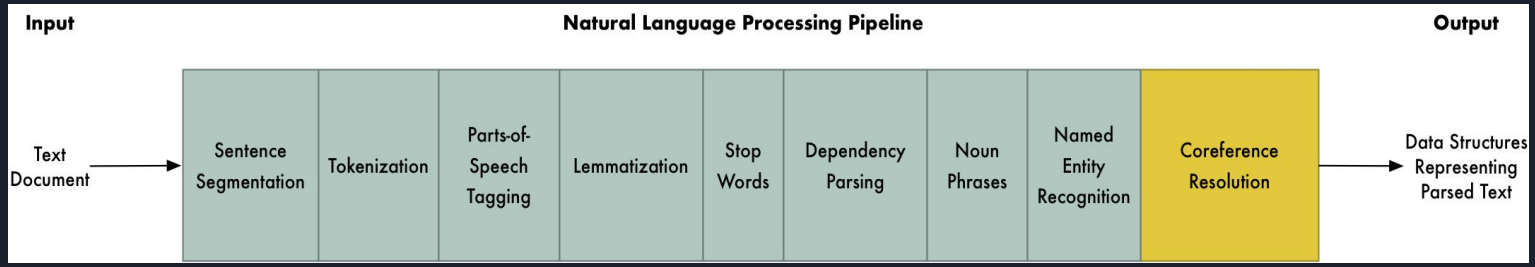
- Provide Data Science related answers through Conversational Mode of Chatbot.
- Implementing Forms for storing the details of people seeking Internship with Sabudh.
- Implementing Voice assistance for chatbot.

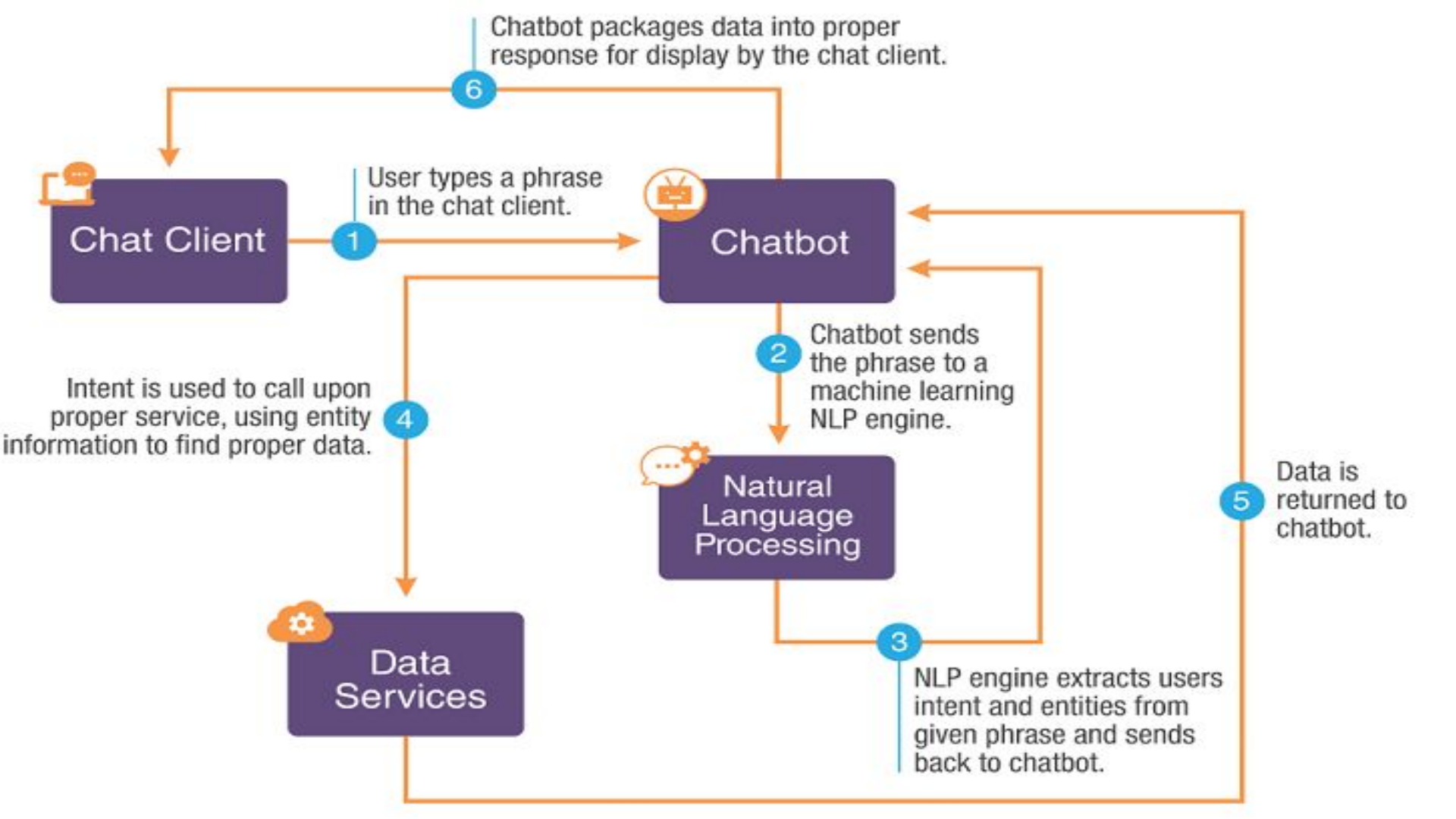


Technologies Used

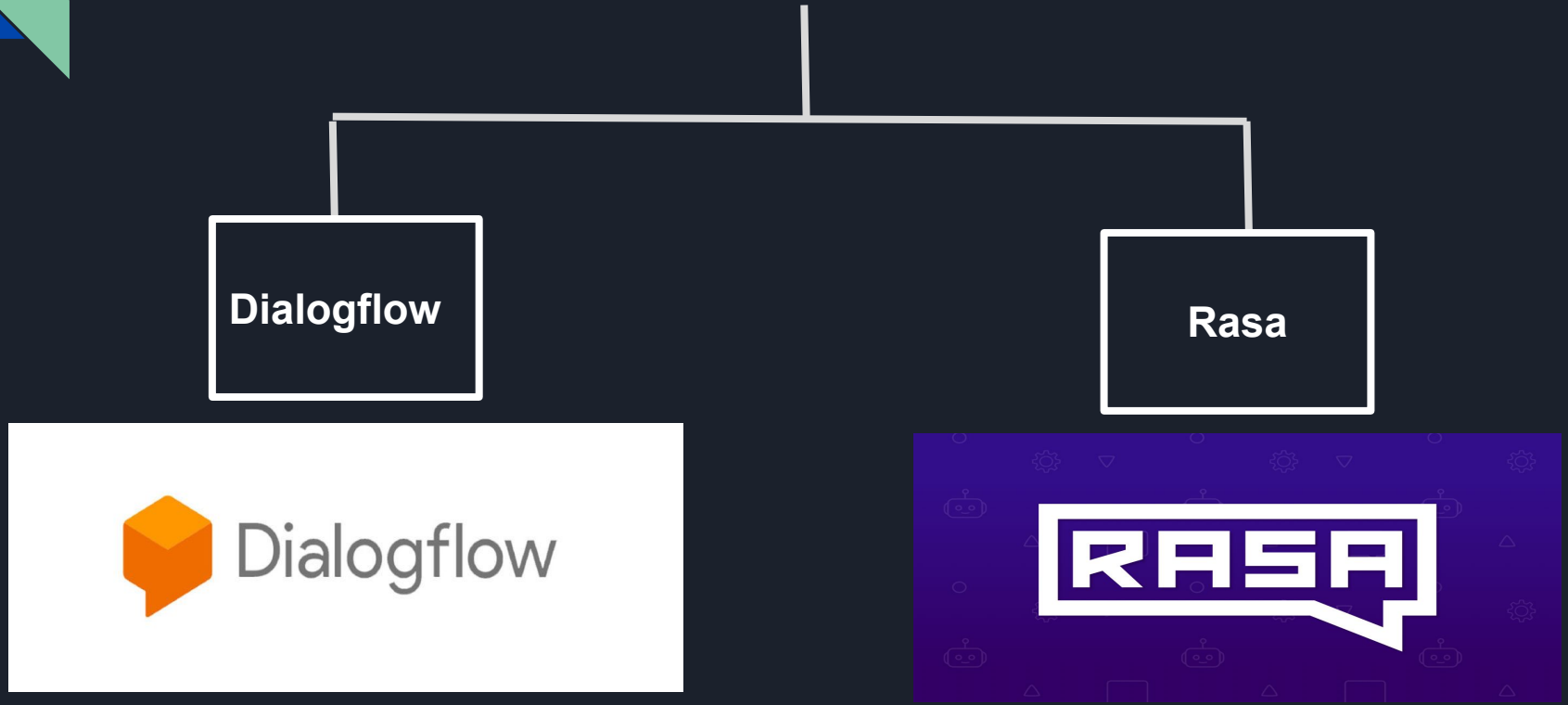
NLP (Natural Language Processing)

It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages. It helps developers to organize knowledge for performing tasks such as *translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction, and topic segmentation*.




Chatbot packages data into proper response for display by the chat client.
6
Chat Client
User types a phrase in the chat client.
1
Chatbot
2
Chatbot sends the phrase to a machine learning NLP engine.
Intent is used to call upon proper service, using entity information to find proper data.
4
Natural Language Processing
5
Data is returned to chatbot.
Data Services
3
NLP engine extracts users intent and entities from given phrase and sends back to chatbot.

Frameworks for a ChatBot

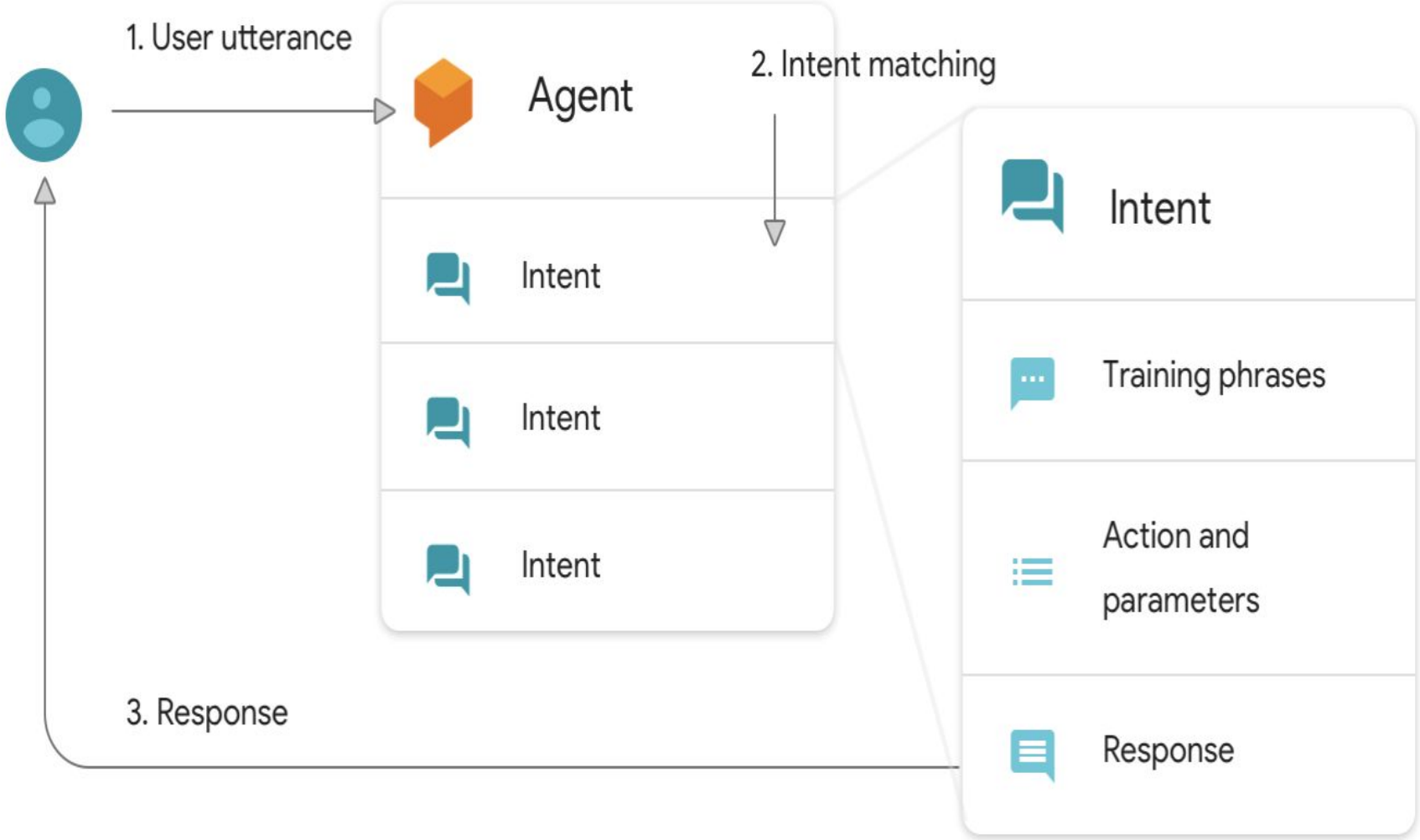


Dialogflow

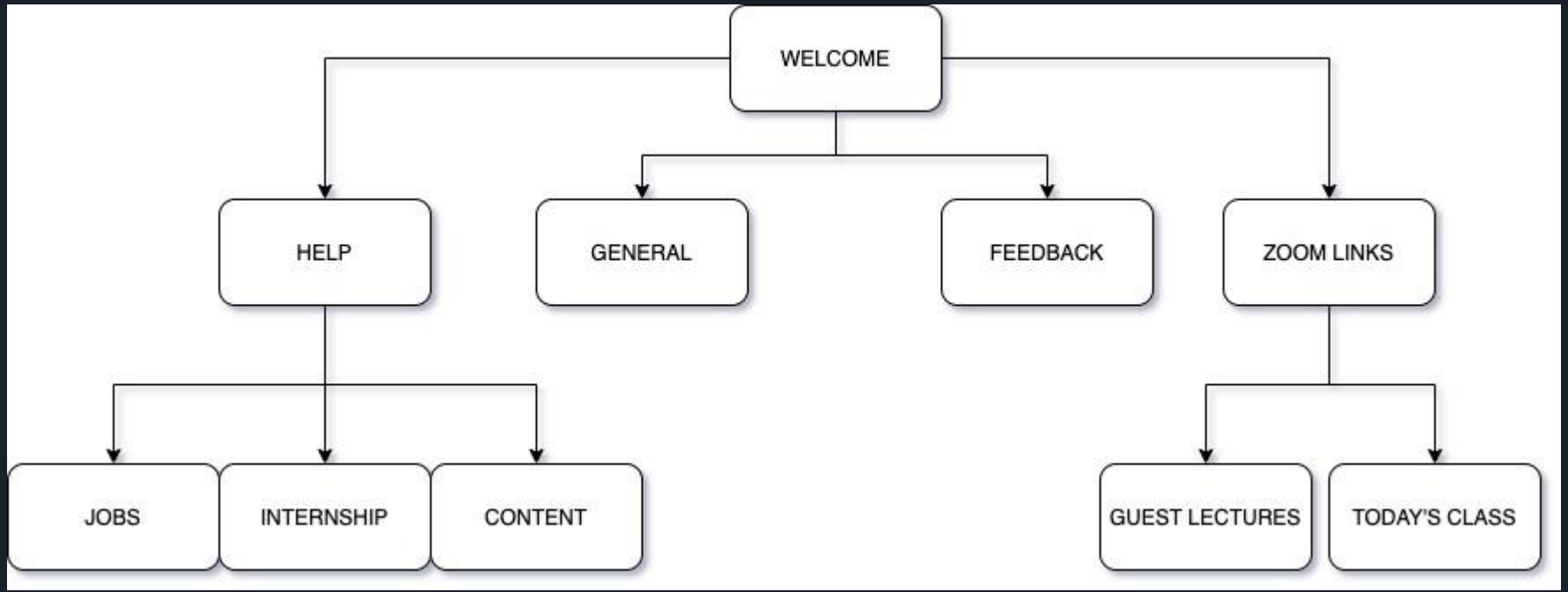
Dialogflow is backed by Google and is the most widely used tool to build Actions for more than 400 M+ Google Assistant devices.

ADVANTAGES OF DIALOGFLOW

- Google Speech — one of the best Natural Language Processing(NLP) systems offered by Google.
- Robust context based input recognition.
- One bot can be integrated on multiple chat/messenger platforms.
- Not only chat/messenger platforms, it also supports wearables and devices like Google Home and Amazon Alexa.
- Free for normal use. The pricing depends upon the usage of the app as the usage increases. (more details below)
- Multiple languages are supported. Currently supported languages are English, Danish, Dutch, French, German, Hindi, Indonesian, Italian, Japanese, Korean, Norwegian, Portuguese (Brazil), Russian, Spanish, Swedish, Thai.


1. User utterance
Agent
2. Intent matching
Intent
Intent
Intent
Intent
Training phrases
Action and parameters
Response
3. Response

Intents in Dialogflow



Disadvantages or why we shifted to RASA

- Creating mesh was not possible in Dialog Flow or in other way difficult to control the flow of the communication.
- Connecting with database such as firebase or mysql was not free.
- You can only provide one webhook for each project. This essentially means that the entire chatbot must have exactly one webhook instead of choosing multiple webhooks on an intent-by-intent basis.

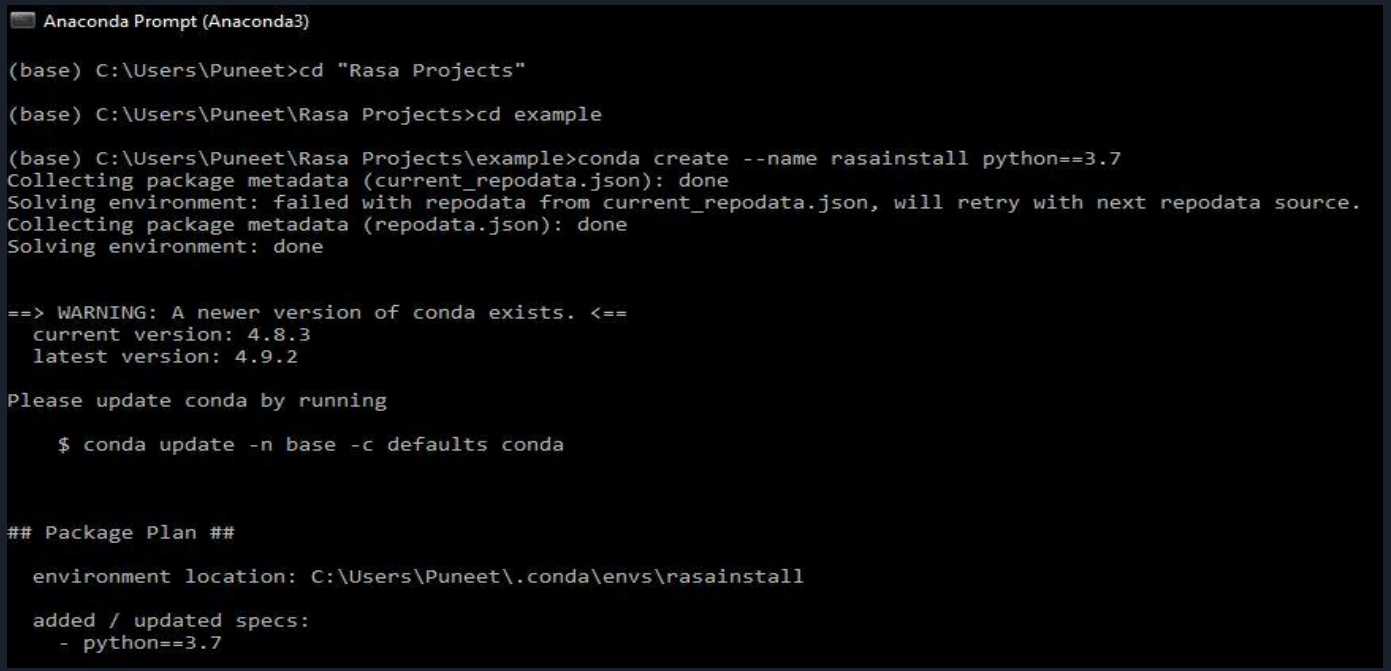
Rasa is an open source framework. It has two major components Rasa NLU and Rasa Core. Rasa NLU is responsible for natural language understanding. Rasa core is a framework for building conversational chatbot. Rasa core allows more sophisticated dialogue, trained using interactive and supervised machine learning.

1. **Ease of Access-** With Rasa, you do not need a specialized hardware, GPU's etc.
2. **Hiding Complexity-** As Rasa grows in functionality, the user environment actually becomes more simplistic and intuitive.
3. **Conversation driven development**
4. **Dialog management-** Rasa uses machine learning to learn conversational patterns and predict response; based on the context etc.
5. **Ease of configuration and change of pipeline -** You can fully customize your NLU pipeline by combining components in the *config.yml* file.

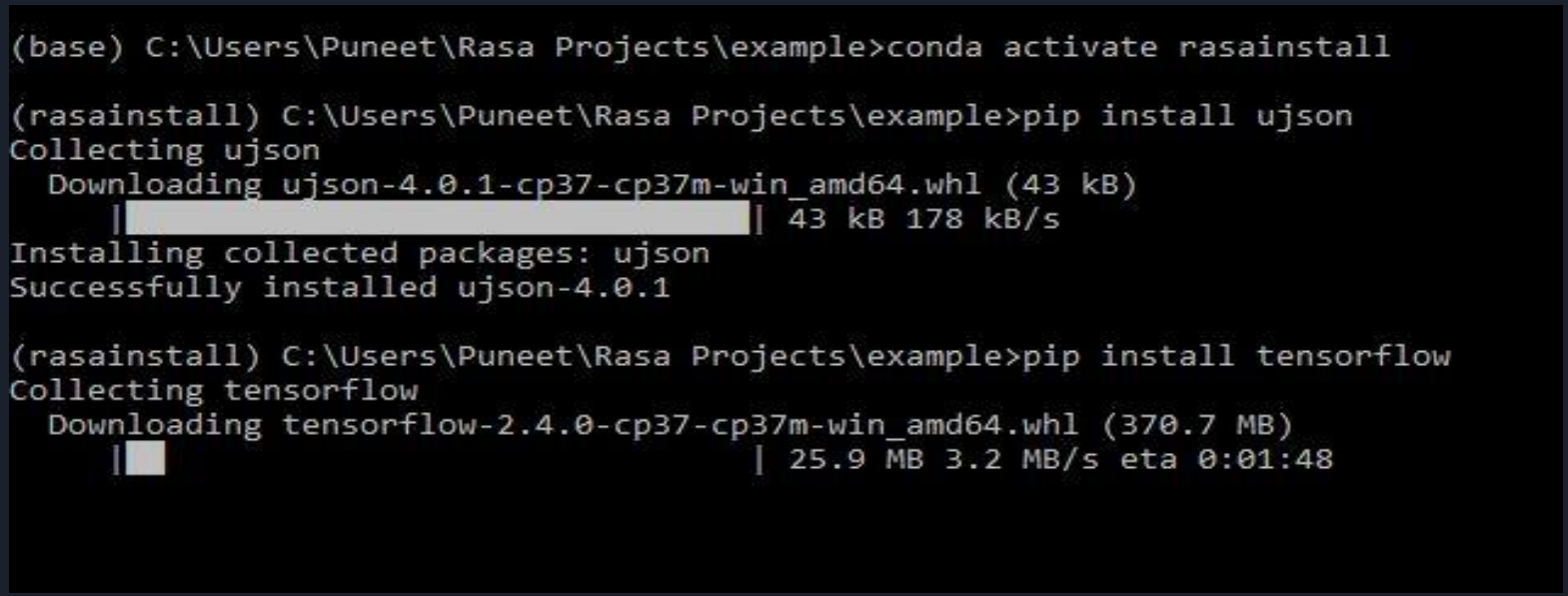
Workflow of RASA

RASA Installation

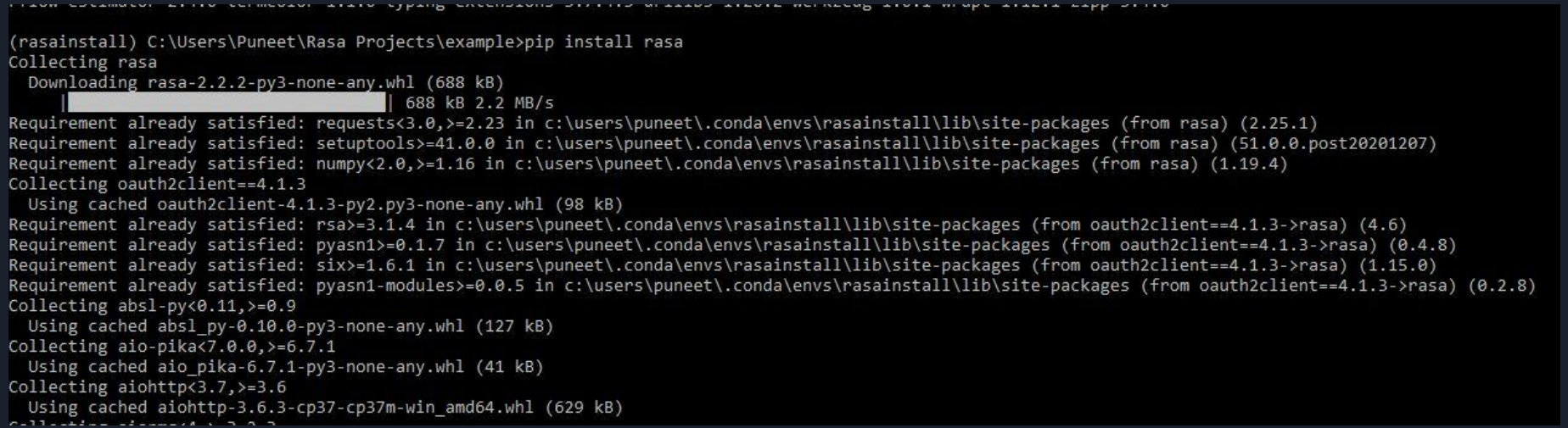
You can install Rasa Open Source using pip (requires Python 3.6, 3.7 or 3.8). In Anaconda terminal, prior installation, move to the folder where you want to install rasa. Give name to your rasa environment and specify version.



Activate the environment and pip install the ujson and tensorflow package which are required for the smooth working of rasa.



The command used is given by ***pip3 install rasa.*** Along with Rasa we also need to install ***vs build tools*.**

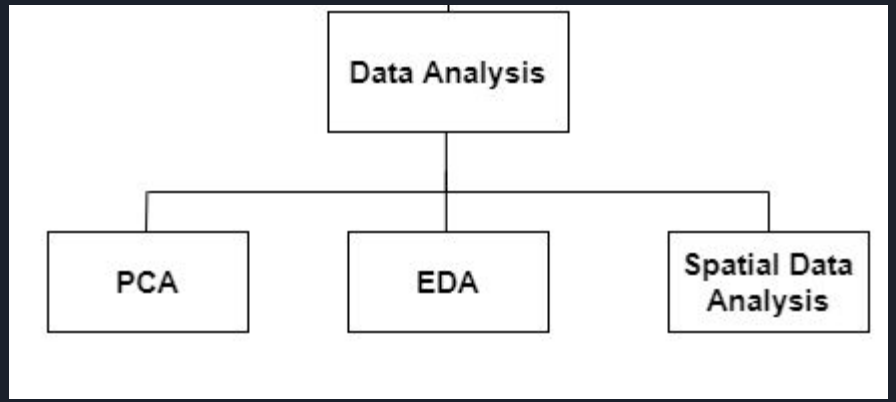


Flow of Intents in Rasa Chatbot

We have considered mainly 4 groups under Data Science:-

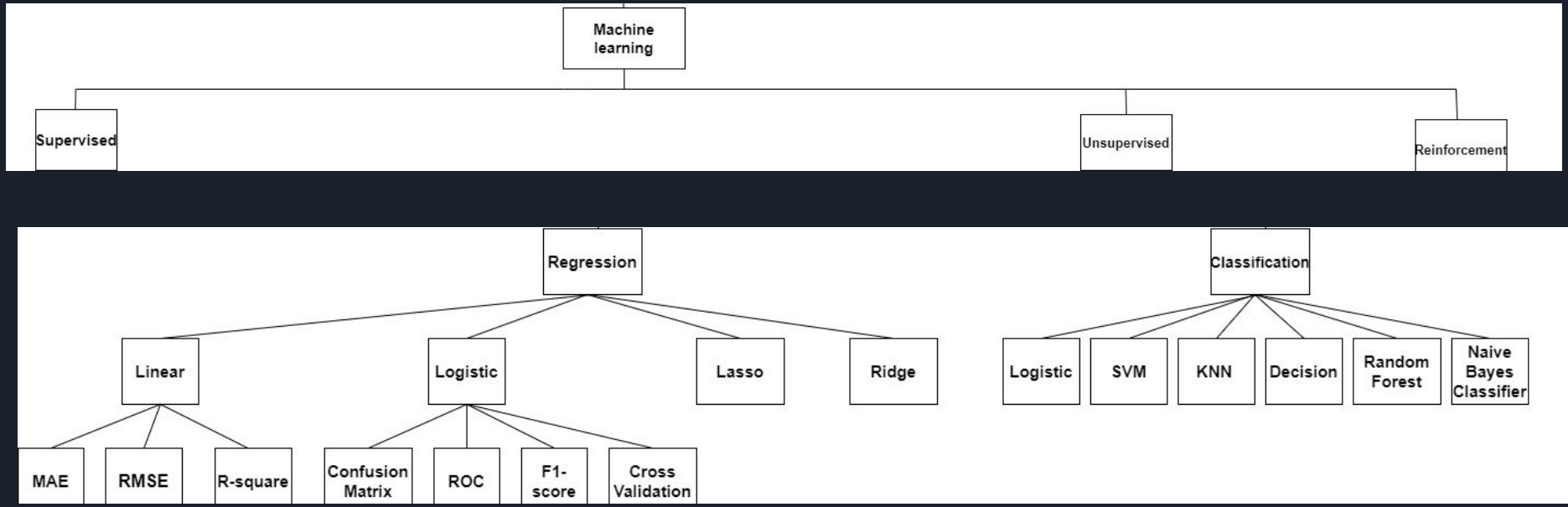
1. Data Analysis

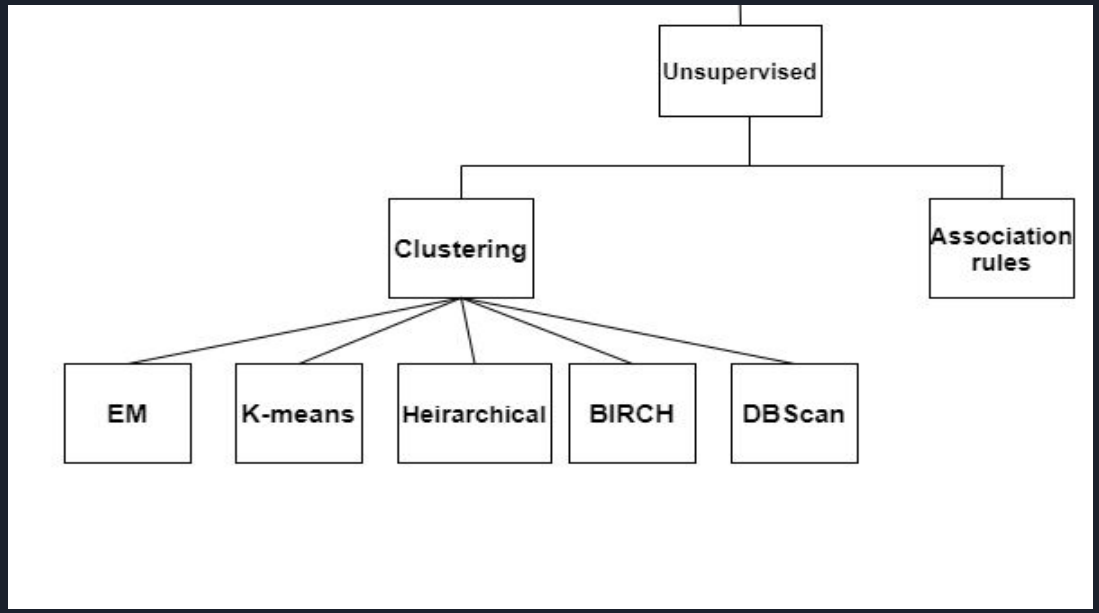
   Data Analysis is further sub grouped into three parts.

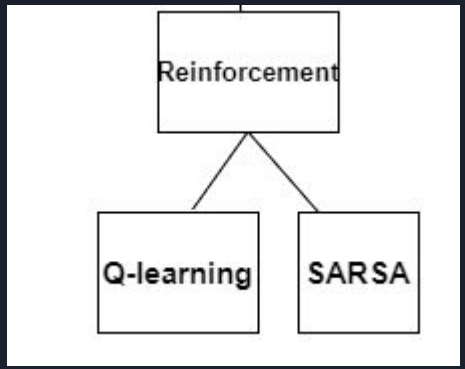


2. Machine Learning

Machine Learning is further subdivided into Supervised, Unsupervised and Reinforcement Learning.
Supervised learning is further divided into Regression and Classification.
Unsupervised learning is further divided into Classification and Association Rules.
Similarly, Reinforcement learning is further divided into Q-learning and SARSA.

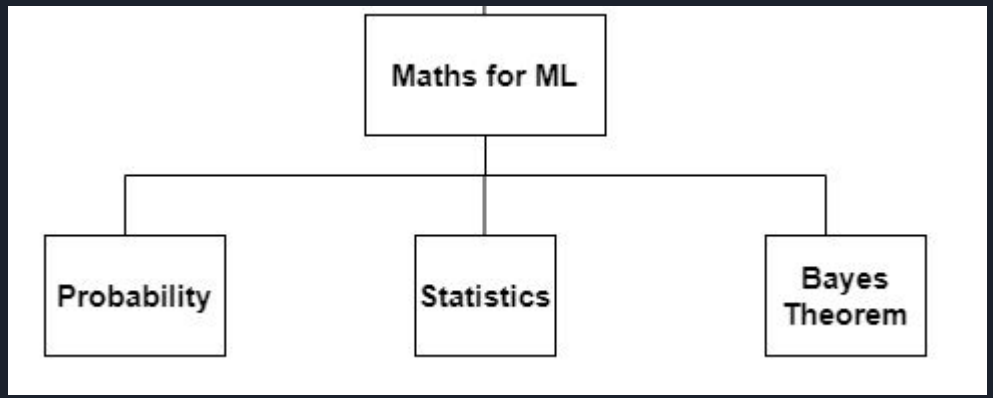



Unsupervised
Clustering
Association rules
EM
K-means
Heirarchical
BIRCH
DBScan


Reinforcement
Q-learning
SARSA

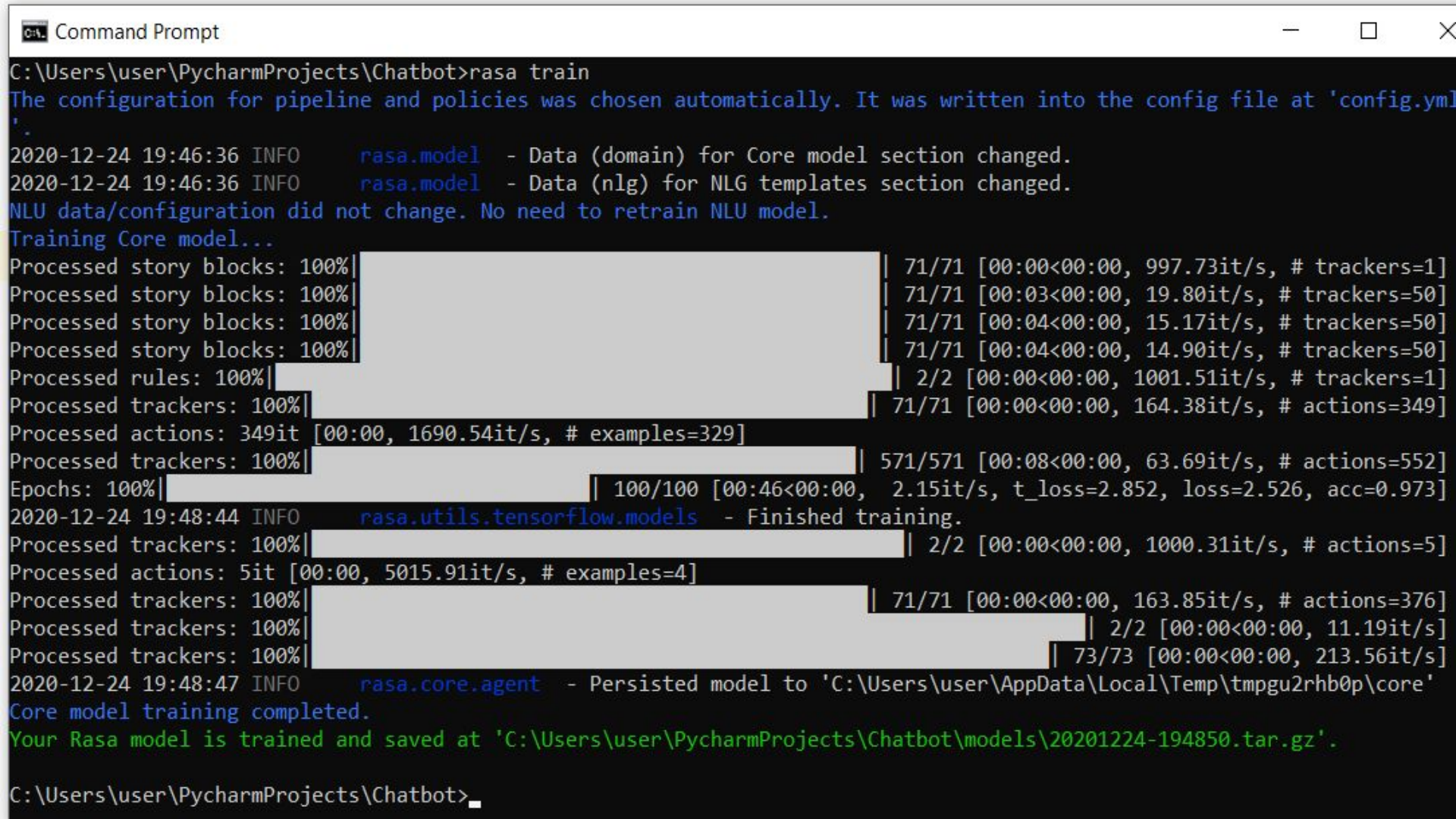
3. Maths for Machine Learning

It is further divided into three subparts- Probability, Bayes theorem and Statistics



4. Data Visualisation

```yaml
language: en
# configuration pipeline
pipeline:
   - name: WhitespaceTokenizer
   - name: RegexFeaturizer
   - name: LexicalSyntacticFeaturizer
   - name: CountVectorsFeaturizer
   - name: CountVectorsFeaturizer
     analyzer: char_wb
     min_ngram: 1
     max_ngram: 4
   - name: DIETClassifier
     epochs: 100
   - name: EntitySynonymMapper
   - name: ResponseSelector
     epochs: 100
   - name: FallbackClassifier
     threshold: 0.5
     ambiguity_threshold: 0.1

# Configuration for Rasa Core.
# https://rasa.com/docs/rasa/core/policies/

# Configuration Policies
policies:
   - name: MemoizationPolicy
   - name: TEDPolicy
     max_history: 5
     epochs: 100
```

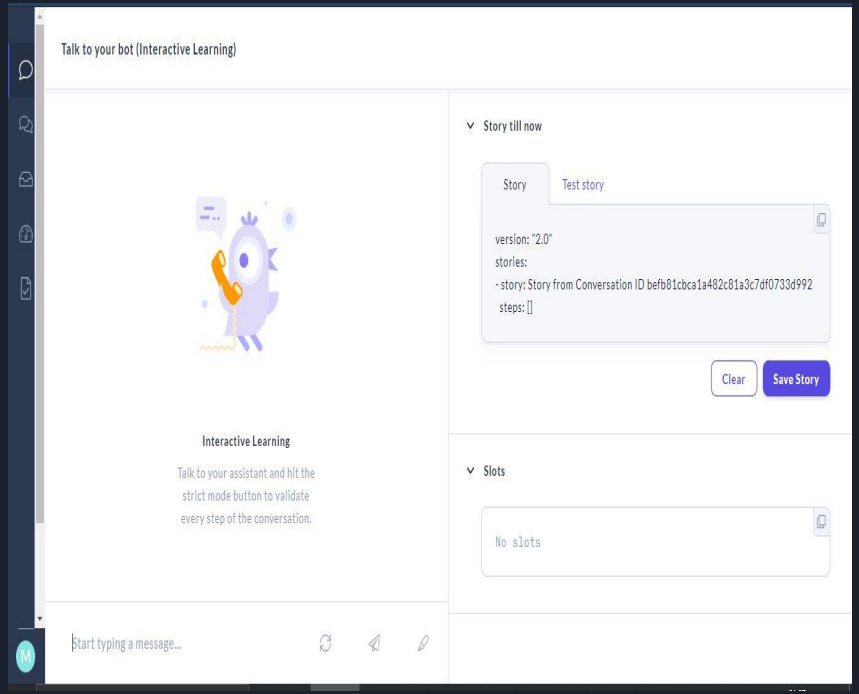

Command Prompt
C:\Users\user\PycharmProjects\Chatbot>rasa train
The configuration for pipeline and policies was chosen automatically. It was written into the config file at 'config.yml'.
2020-12-24 19:46:36 INFO rasa.model - Data (domain) for Core model section changed.
2020-12-24 19:46:36 INFO rasa.model - Data (nlg) for NLG templates section changed.
NLU data/configuration did not change. No need to retrain NLU model.
Training Core model...
Processed story blocks: 100%| | 71/71 [00:00<00:00, 997.73it/s, # trackers=1]
Processed story blocks: 100%| | 71/71 [00:03<00:00, 19.80it/s, # trackers=50]
Processed story blocks: 100%| | 71/71 [00:04<00:00, 15.17it/s, # trackers=50]
Processed story blocks: 100%| | 71/71 [00:04<00:00, 14.90it/s, # trackers=50]
Processed rules: 100%| | 2/2 [00:00<00:00, 1001.51it/s, # trackers=1]
Processed trackers: 100%| | 71/71 [00:00<00:00, 164.38it/s, # actions=349]
Processed actions: 349it [00:00, 1690.54it/s, # examples=329]
Processed trackers: 100%| | 571/571 [00:08<00:00, 63.69it/s, # actions=552]
Epochs: 100%| | 100/100 [00:46<00:00, 2.15it/s, t_loss=2.852, loss=2.526, acc=0.973]
2020-12-24 19:48:44 INFO rasa.utils.tensorflow.models - Finished training.
Processed trackers: 100%| | 2/2 [00:00<00:00, 1000.31it/s, # actions=5]
Processed actions: 5it [00:00, 5015.91it/s, # examples=4]
Processed trackers: 100%| | 71/71 [00:00<00:00, 163.85it/s, # actions=376]
Processed trackers: 100%| | 2/2 [00:00<00:00, 11.19it/s]
Processed trackers: 100%| | 73/73 [00:00<00:00, 213.56it/s]
2020-12-24 19:48:47 INFO rasa.core.agent - Persisted model to 'C:\Users\user\AppData\Local\Temp\tmpgu2rhb0p\core'
Core model training completed.
Your Rasa model is trained and saved at 'C:\Users\user\PycharmProjects\Chatbot\models\20201224-194850.tar.gz'.
C:\Users\user\PycharmProjects\Chatbot>

Command Prompt - rasa shell

```
C:\Users\user\PycharmProjects\Chatbot>rasa shell
2020-12-24 19:51:36 INFO     rasa.model  - Loading model models\20201224-194850.tar.gz...
2020-12-24 19:51:41 INFO     root  - Connecting to channel 'cmdline' which was specified by the '--connector' argument.
Any other channels will be ignored. To connect to all given channels, omit the '--connector' argument.
2020-12-24 19:51:41 INFO     root  - Starting Rasa server on http://localhost:5005
2020-12-24 19:51:41 INFO     rasa.model  - Loading model models\20201224-194850.tar.gz...
2020-12-24 19:52:08 INFO     root  - Rasa server is up and running.
Bot loaded. Type a message and press enter (use '/stop' to exit):
Your input ->  hi
Hey! I am a Data Science bot. Data science is a concept to unify statistics, data analysis and their related methods in
order to understand and analyze actual phenomena with data. It uses techniques and theories drawn from many fields withi
n the context of mathematics, statistics, computer science, domain knowledge and information science. Please select a Ca
tegory to continue. Data Analysis / Mathematics for Machine Learning / Machine Learning / Data Visualization
Your input ->  Machine Learning
It is a lazy learning algorithm that stores all instances corresponding to training data in n-dimensional space. It is a
 lazy learning algorithm as it does not focus on constructing a general internal model, instead, it works on storing ins
tances of training data. Classification is computed from a simple majority vote of the k nearest neighbors of each point
. It is supervised and takes a bunch of labeled points and uses them to label other points. To label a new point, it loo
ks at the labeled points closest to that new point also known as its nearest neighbors. It has those neighbors vote, so
whichever label the most of the neighbors have is the label for the new point. The "k" is the number of neighbors it che
cks. This algorithm is quite simple in its implementation and is robust to noisy training data. Even if the training dat
a is large, it is quite efficient. The only disadvantage with the KNN algorithm is that there is no need to determine th
e value of K and computation cost is pretty high compared to other algorithms. For more information, visit the link belo
w:- https://www.edureka.co/blog/classification-in-machine-learning/
Do you want to know about the Classification techniques? I can help you with Logistic / SVM / KNN / Decision trees / Ran
dom Forest / Naive Bayes Classifier. Choose any category to continue.
Your input ->
```

Rasa X or Rasa UI

Rasa has taken a very innovative and unique approach to continuous improvement of chatbots. The approach is Conversation Driven Development tool

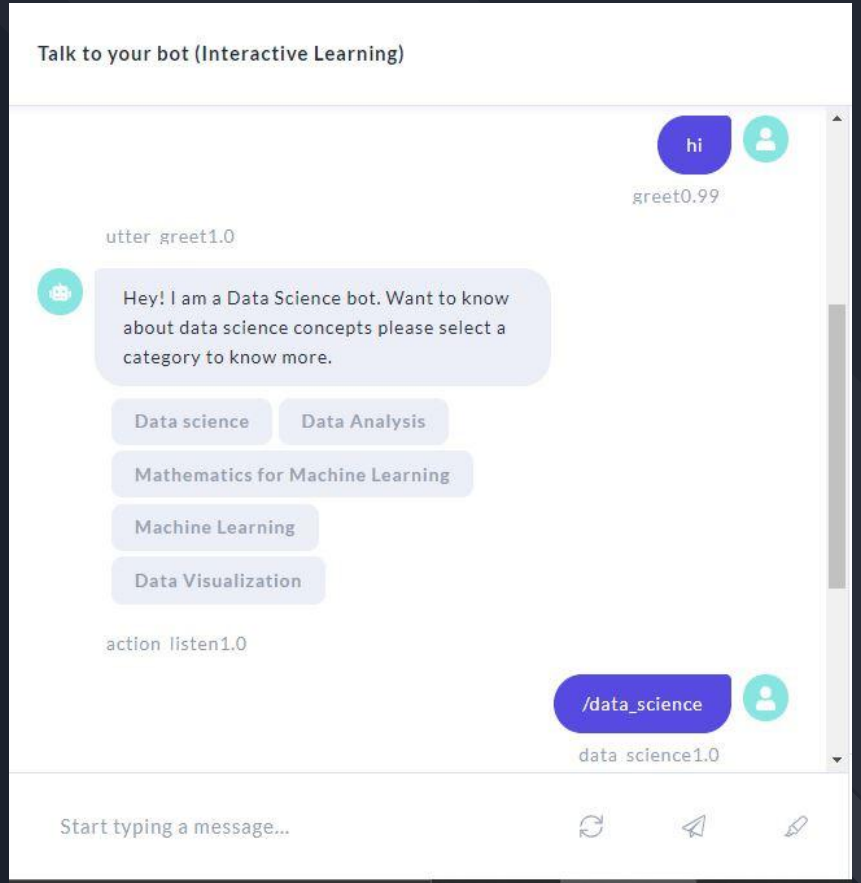


Terminal: Local (4) × Local × +

```
(c) 2020 Microsoft Corporation. All rights reserved.

(chatbot) C:\Users\Puneet\Rasa Projects\chatbot_remote>conda activate installingrasa

(installingrasa) C:\Users\Puneet\Rasa Projects\chatbot_remote>rasa train
The configuration for pipeline and policies was chosen automatically. It was written into the config file at 'config.yml'.
2020-12-24 20:36:35 INFO     rasa.shared.utils.validation  - The 'version' key is missing in the training data file C:\Users\Puneet\Rasa Projects\chatbot_remote\domain.yml. Rasa Open Source
 will read the file as a version '2.0' file. See https://rasa.com/docs/rasa/training-data-format.
Nothing changed. You can use the old model stored at 'C:\Users\Puneet\Rasa Projects\chatbot_remote\models\20201222-105418.tar.gz'.

(installingrasa) C:\Users\Puneet\Rasa Projects\chatbot_remote>rasa x
Starting Rasa X in local mode...
c:\users\puneet\.conda\envs\installingrasa\lib\site-packages\rasa\shared\utils\io.py:93: UserWarning: Issue found in 'C:\Users\Puneet\AppData\Local\Temp\tmphysyt983':
Found intent 'mood_great' in stories which is not part of the domain.
  More info at https://rasa.com/docs/rasa/stories
c:\users\puneet\.conda\envs\installingrasa\lib\site-packages\rasa\shared\utils\io.py:93: UserWarning: Issue found in 'C:\Users\Puneet\AppData\Local\Temp\tmphysyt983':
Found intent 'mood_unhappy' in stories which is not part of the domain.
  More info at https://rasa.com/docs/rasa/stories
c:\users\puneet\.conda\envs\installingrasa\lib\site-packages\rasa\shared\utils\io.py:93: UserWarning: Issue found in 'C:\Users\Puneet\AppData\Local\Temp\tmphysyt983':
Found intent 'bot_challenge' in stories which is not part of the domain.
  More info at https://rasa.com/docs/rasa/stories
[2020-12-24 21:02:04 +0530] [15212] [INFO] Goin' Fast @ http://0.0.0.0:5002
2020-12-24 21:02:04 INFO     sanic.root  - Goin' Fast @ http://0.0.0.0:5002

The server is running at http://localhost:5002/login?username=me&password=KY69VFZptNk9

[2020-12-24 21:02:05 +0530] [15212] [INFO] Starting worker [15212]
2020-12-24 21:02:05 INFO     sanic.root  - Starting worker [15212]
```

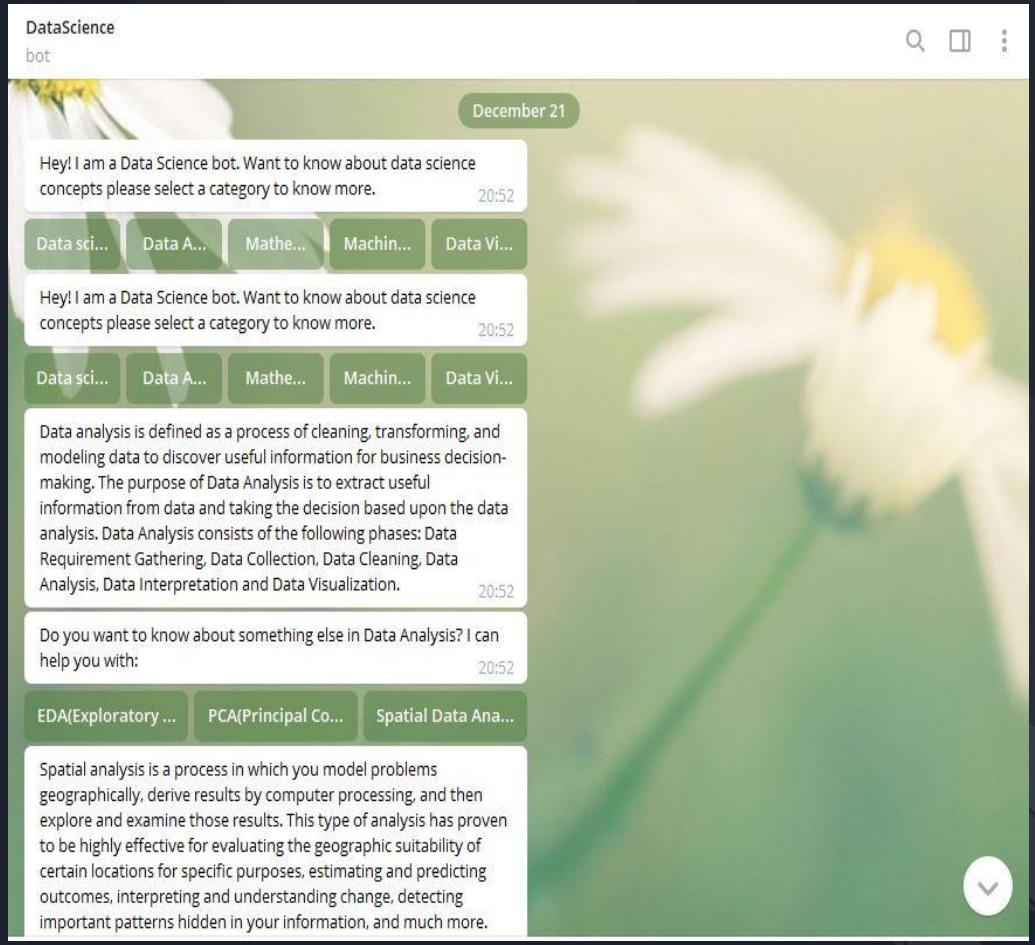
Talk to your Bot

In rasa x, you can have content driven conversation according to the model.



Integration with telegram

You can converse on with data science bot on telegram as well.



Integration with Voice assistant

You can give commands to your bot with your voice and it returns the response as text.

```
Speak Anything :
you said : hello
Sending message now.....
Bot:  Hey! I am a Data Science bot. Want to know about data science concepts please select a category to know more.
Speak Anything :
you said : machine learning
Sending message now.....
```

```
2020-12-28 06:12:43 INFO     rasa.core.test  - Evaluation Results on END-TO-END level:
2020-12-28 06:12:43 INFO     rasa.core.test  -  Correct:          5 / 6
2020-12-28 06:12:43 INFO     rasa.core.test  -  F1-Score:         0.909
2020-12-28 06:12:43 INFO     rasa.core.test  -  Precision:        1.000
2020-12-28 06:12:43 INFO     rasa.core.test  -  Accuracy:         0.833
2020-12-28 06:12:43 INFO     rasa.core.test  -  In-data fraction: 0.867
2020-12-28 06:12:43 INFO     rasa.core.test  - Stories report saved to results\story_report.json.
2020-12-28 06:12:44 INFO     rasa.core.test  - Evaluation Results on ACTION level:
2020-12-28 06:12:44 INFO     rasa.core.test  -  Correct:          28 / 30
2020-12-28 06:12:44 INFO     rasa.core.test  -  F1-Score:         0.964
2020-12-28 06:12:44 INFO     rasa.core.test  -  Precision:        0.984
2020-12-28 06:12:44 INFO     rasa.core.test  -  Accuracy:         0.953
2020-12-28 06:12:44 INFO     rasa.core.test  -  In-data fraction: 0.867
2020-12-28 06:12:48 INFO     rasa.utils.plotting  - Confusion matrix, without normalization:
[[ 0  0  0  0  0  0  0  0  0  0]
 [ 0 13  0  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0]
 [ 0  0  0  2  0  0  0  0  0  0]
 [ 0  0  0  0  1  0  0  0  0  0]
```