REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Edubot: RACE Powered Chatbot

Pradeep Thota

SRN: R19MBA63

Date: 20/08/2022

MBA in Business Analytics

Capstone Project Presentation
Year: I

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Agenda

01 Introduction
Back Ground | Current status | Why this study

02 Literature Review
Seminal works | Summary | Research Gap

03 Problem Statement
Business Problem | Analytics Solution

04 Project Objectives
Primary & Secondary Objectives | Expected Outcome

05 Project Methodology
Conceptual Framework | Research Design

06 Business Understanding
Business Context | Monetary Impact

07 Data Understanding
Data Collection | Variables

08 Data Preparation
Pre-processing | Process | Techniques

09 Descriptive Analytics
Univariate | Bivariate | Hypothesis

10 Modeling
Machine Learning | Model Evaluation | Insights

11 Model Deployment
Applications | Demo

12 Suggestions and Conclusions
Insights | Next Step | Future Scope

13 Annexure
References | Publications | Plagiarism Score

Introduction

Background | Current status | Why this study

Chatbots are defined as computer programs that replicate human-like conversations by using linguistic communication structures.

The term “Chatbots” is derived from major words that are “chat” in lieu of the conversational attributes and “bot” for the robot

There are mainly 2 types of chatbot

1) Chatbot – Text-based
2) Voice bot – Image-based

In addition to this, there are 2 more chatbots that are still in development with the full version

1) Image bot – Image-based
2) Video bot – Video-based

Literature Review

Seminal works | Summary | Research Gap

- Have reviewed a minimum of 15 research papers
- There are 6 types of parameters they are:
    1. Open and closed domain names in the information domain
    2. Chatbots that are interpersonal, intrapersonal, and inter-agent
    3. Informative, entirely chat-based objectives.
    4. Task-based fully enter processing and reaction generation: rule-based full version, retrieval-based full version, and generative full version
    5. Resources useful for Humans
    6. Create either open-supply or closed systems.

- As technology is getting advances most chatbots are built using AI.
- With AI's help, we can see everyone can build chatbots with no code using IBM Watson, Amazon Lex, Google dialog flow chatbots, etc.
- After building the chatbots we can deploy them on many platforms as per business requirements like MIM like WhatsApp, Telegram, etc., Web applications, Social media, etc.

Problem Statement

Business Problem | Analytics Solution

- REVA Academy of Corporate Excellence, REVA University's existing system has a chat interface where an executive will be chatting from the backend.
- FAQs alone cannot address all the concerns of the user so an interactive and user-friendly way of conversation is required that can have the least waiting time with accurate and suitable answers to the user query.
- In the currently existing process when a customer needs any immediate assistance from the university, the user can chat through a chat window available on the official website, in this process one of the executives will connect to assets the customer so that all the questions can be answered.
- But the problem arises in non-working hours when no executive can be connected to the chat and will be very hard to get assistance for hours till the executive connects to the chat.

Project Objectives

Primary & Secondary Objectives | Expected Outcome

- The objective of this project is to develop an Interactive chatbot that will be assisting the customers 24/7 round the clock with their multiple questions related to different courses available in RACE like Master of Business Administration in Business Analytics, Master of Technologies in Cybersecurity/ Artificial Intelligence, etc., with all the details like duration, fee, eligibility, affiliation, etc., without delay in the response so going to avoid the long queue of waiting for customers to get a response.

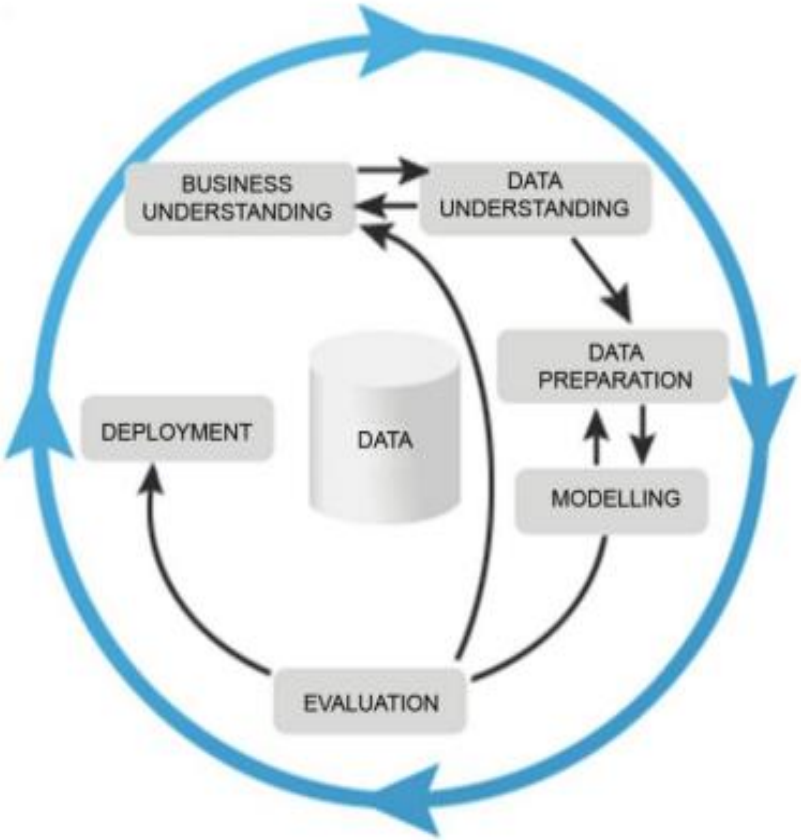
- Data collection is the crucial part of the project and have collected the dump of it from the previous chat history available on the server and prepared it into a CSV file and used it as the input for this project. The collected data are parsed, and information is extracted using different techniques in NLP.

- The instant output/response will be displayed to the user as soon as the user raises a question/doubt in the chat window once the model is implemented and deployed.

Project Methodology

Conceptual Framework | Research Design

- **Business Understanding** — The goal of this stage is to understand the business goal and then convert it into a measurable and specific project goal and then formalize it as a problem statement.
- **Data Understanding** — The goal of this stage is to gather data and then explore and comprehend the data.
- **Data Preparation** – The goal of this stage is to select the final data which will be relevant to the data mining objectives, and clean and transform the data.
- **Data Modelling** - The goal of this stage is, to apply the modeling techniques and record them.
- **Model Evaluation** – The goal of this stage is, to assess the degree to which the model meets the business requirements and to test the model in real applications.
- **Deployment** - The goal of this stage is to determine the model deployment strategy based on evaluation results and plan for monitoring and maintenance of models in the business environment.



Business Understanding

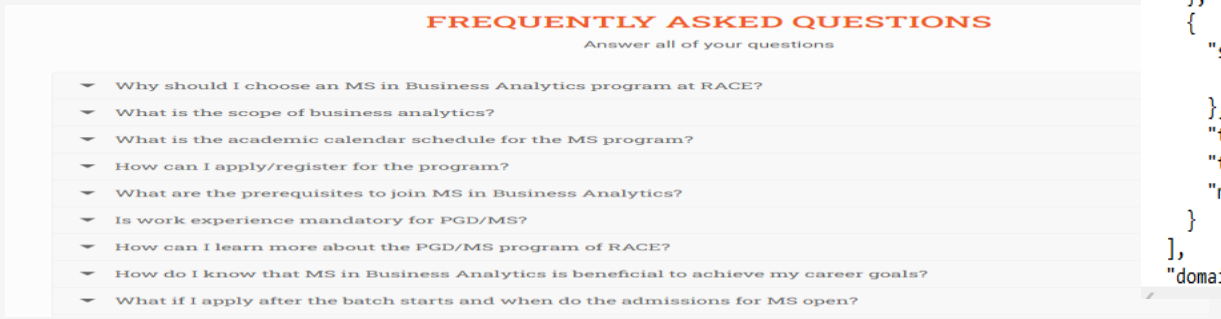
Business Impact | Challenges | Monetary Impact

- REVA Academy of Corporate Excellence, REVA University needs transformation from traditional chat with an executive to support that helps the customer to get an instant response to all their questions.
- By implementing this the business need not assign a dedicated resource to solve all the questions raised by the customer related to the course.
- By executing this business can save time and gets monetary gain which is spent on the resource who is working on the live chat session and providing the exact information that the customer needs at the same time without any delay.

Data Understanding

Data Collection | Variables

- Data comprises unstructured data of chat which are in the form of JSON, excel, and notepad.
- Data is collected from the university GitHub repository and all the FAQs which are available on RACE REVA's official website (https://race.reva.edu.in/pgdm-mba-in-business-analytics/#faq).
- Here data is unstructured with many special characters, date, time, executive name, etc, and FAQs are in notepad as downloaded from the website in unstructured text format and are converted into structured text format of CSV file.




```
{
  "id": "206ef860-4c20-11eb-9856-b5a88d1782cb",
  "type": "chat",
  "pageId": "5bed1e510e6b3311cb795470",
  "visitor": {
    "name": "admin",
    "id": "4170ebd01f43b7487217d4f4ce40e8771f498f935631dc5860f6b83cb168cec7",
    "email": "vijay.kumawat@innoserv.co.in"
  },
  "location": {
    "countryCode": "IN",
    "city": "Vijayawada"
  },
  "messageCount": 1,
  "chatDuration": 0,
  "rating": 0,
  "createdOn": "2021-01-01T10:57:36.735Z",
  "messages": [
    {
      "sender": {
        "t": "s"
      },
      "type": "msg",
      "time": "2021-01-01T10:57:36.735Z",
      "msg": "Welcome to RACE, if you need help simply reply to this message, we are online and ready to help!"
    },
    {
      "sender": {
        "t": "v"
      },
      "type": "msg",
      "time": "2021-01-01T10:59:12.233Z",
      "msg": "hi"
    }
  ],
  "domain": "race.reva.edu.in"
```

Data Preparation

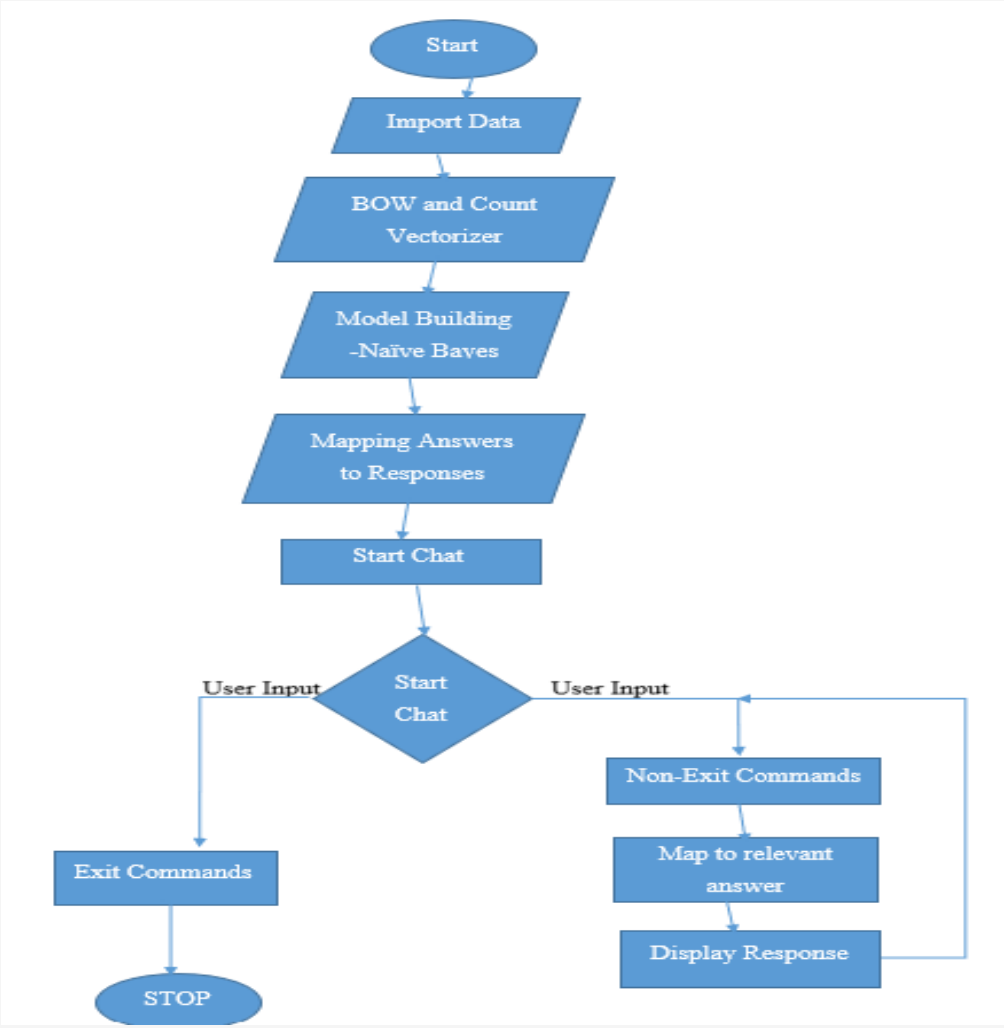
- The input for Chatbot is .CSV file and all the data for this CSV are collected from JSON, excel, and notepad which are downloaded from the RACE Git repository and from FAQs of the RACE official website.
- CSV file contains 2 columns heading with Questions and Responses, all the questions which are captured manually from the raw data file has put into CSV file of Questions column and for this question have provided with the numerical same number for all the relevant question and its mapping is done in model preparation step.

	Question	response
2	i need to discuss on the cybersecurity program	1
3	PG Diploma/M.Tech/MS in\nCYBERSECURITY	1
4	For dip in cyber ersececurity	1
5	Hello, I would like to know more about MS in Cyber Security at your Univers	1
6	I want to do mtech in cybersecurity	1
7	I am looking for PG Diploma in cybersecurity	1
8	Do u provide training in cybersecurity	1
9	Looking for Cybersecurity courses	1
10	How much is the course fees for cyber security	4
11	Near information about cyber security diploma course	1
12	I heard that REVA uni will be offering an MS program in AI in collaboration v	1
13	I would like to know M tech cyber security program	1
14	I want to know about data analyst course	1
15	admission m.tech	3
16	need some details about mba in business analytics	2
17	Am 33 years old I can do MBA course	2
18	class for mba	2
19	I wanted to do M-Tech	3

Modeling

- Once after importing the necessary packages and corpus file then all the data is taken into Bag of Words and then converted into Count Vectorizer to transform the data contains in the Questions column in the corpus into the frequency(count) of the words in the text.
- As the data preparation step is done then have built the model based on Naïve Bayes – MultinominalNB as the classifier for the model, as the importing is done then have taken all the count vectorizer corpus as the training vectors.
- As the next step of data modeling, we have all the questions categorized and provided with numbers for each category in the raw corpus file, so must map all the numbers to the relevant answers within the code. In this process, we have a set of exit commands that can break the chain of this chat if the user provides a response with any of these exit commands in the chat window.

Modeling


Start
Import Data
BOW and Count Vectorizer
Model Building -Naïve Bayes
Mapping Answers to Responses
Start Chat
Start Chat
User Input
User Input
Exit Commands
STOP
Non-Exit Commands
Map to relevant answer
Display Response

Model Evaluation

Results | Interpretation | Insights

- In this project we have calculated the accuracy for each and every question.
- Accuracy for each question is calculated = No. of points classified correctly / all points (in the test set)
- Accuracy of the entire model is based on multinomialNB classification and have tested with a minimum of 20 chats consisting of 10 questions in each chat and got an accuracy average of 78.3%.

Model Deployment

Demonstration

- After performing data Processing and multiple iterations to get the approval of the client then we can deploy this on their official website and this model become the central heart of all the question of candidate who needs information about the course through this chatbot and has provided this entire model in .py file so it can deploy the same in a different flavor of a platform like WhatsApp, Telegram, Teams, etc. so that user can get information in every platform as per their mode of browse.

Results and Insights

Key Findings | Suggestions

- Data Evaluation results are captured from Chatbot. The dataset used for training the model consists of 96 questions from different sources. Model got trained in Naïve Bayes with classifier as multinomialNB.
- Below are the results consisting of how accurate the model is based on multinomialNB classification and have tested with a minimum of 20 chats consisting of 10 questions in each chat and got an accuracy average of 78.3%.

```
racebot = ChatBot()
racebot.start_chat()
```

```
Hi, I'm RETINA a RACE powered chatbot!!
are you a chat bot
Accurate: 80.53%
I'm RETINA, RACE powered Chatbot.
thank you, I am looking for weekend PG/Mtech  program for AI
Accurate: 77.57%
AI undeniable transformation in the technological world with a PG Diploma/ M. Tech/M.Sc in Artificial Intelligence program offe
red by RACE. The comprehensive learning approach to master the domains of Artificial Intelligence, Data Science, Business Analy
tics, Business Intelligence, and Deep Learning enables the participants to take on challenging roles in the Artificial Intellig
ence domain
What is the fees
Accurate: 99.48%
For Fee detail please contact Program office Abhijit Sinha : +91 95388 74441, Basavaraj Shetty: +91 76250 69676 or write to rac
e@reva.edu.in
bye
Ok, have a great day!
```

Conclusion and Future Work

Proposed solutions | Scope for future work

- In this project have developed a chatbot that is text-based that can help the users to get instant responses to all their questions without any time delay and at the same time it can also save time and cost for the RACE department by not assigning a resource to chat with the users all the time and after implementing this resource can be utilized for another work.
- This project in the future can be enhanced to multiple platforms like Voice bot, Image bot, and Video bot, so users can experience new things which are developed with more advanced technology than normal and that can lead to acquiring more students.

References

Bibliography | Webliography

- Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. *IFIP Advances in Information and Communication Technology*, *584 IFIP*. https://doi.org/10.1007/978-3-030-49186-4_31
- Arruda, D., Marinho, M., Souza, E., & Wanderley, F. (2019). A Chatbot for Goal-Oriented Requirements Modeling. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11622 LNCS*. https://doi.org/10.1007/978-3-030-24305-0_38
- Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: User changing needs and motivations. *Interactions*, *25*(5). https://doi.org/10.1145/3236669
- Brindha, S., Dharan, K. R. D., Samraj, S. J. J., & Nirmal, L. M. (2019). AI based Chatbot for Education Management. *International Journal of Research in Engineering, Science and Management*, *2*(3).
- Brockhus, S., van der Kolk, T. E. C., Koeman, B., & Badke-Schaub, P. G. (2014). The influence of creative self-efficacy on creative performance. *Proceedings of International Design Conference, DESIGN*, *2014-January*.
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2). https://doi.org/10.1037//0033-2909.119.2.197
- Cameron, G., Cameron, D. M., Megaw, G., Bond, R. B., Mulvenna, M., O'Neill, S. B., Armour, C., & McTear, M. (2018). *Back to the Future: Lessons from Knowledge Engineering Methodologies for Chatbot Design and Development*. https://doi.org/10.14236/ewic/hci2018.153
- Chaves, A. P., & Gerosa, M. A. (2021). How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human-Computer Interaction*, *37*(8). https://doi.org/10.1080/10447318.2020.1841438
- Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A ChatBot for Learning Chinese: Learning Achievement and Technology Acceptance. *Journal of Educational Computing Research*, *58*(6). https://doi.org/10.1177/0735633120929622

References

Bibliography | Webliography

- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*. https://doi.org/10.1016/j.future.2018.01.055
- Colace, F., de Santo, M., Lombardi, M., Pascale, F., Pietrosanto, A., & Lemma, S. (2018). Chatbot for e-learning: A case of study. *International Journal of Mechanical Engineering and Robotics Research*, *7*(5). https://doi.org/10.18178/ijmerr.7.5.528-533
- Conde, M., Rodríguez-Sedano, F. J., Hernández-García, Á., Gutiérrez-Fernández, A., & Guerrero-Higueras, Á. M. (2021). Your teammate just sent you a new message! the effects of using telegram on individual acquisition of teamwork competence. *International Journal of Interactive Multimedia and Artificial Intelligence*, *6*(6). https://doi.org/10.9781/ijimai.2021.05.007
- Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research. In *Educational Research* (Vol. 4).
- de Oliveira, J. C., Santos, D. H., & Neto, M. P. (2016). Chatting with Arduino platform through Telegram Bot. *Proceedings of the International Symposium on Consumer Electronics, ISCE*. https://doi.org/10.1109/ISCE.2016.7797406
- Dekker, I., de Jong, E. M., Schippers, M. C., de Bruijn-Smolders, M., Alexiou, A., & Giesbers, B. (2020). Optimizing Students' Mental Health and Academic Performance: AI-Enhanced Life Crafting. In *Frontiers in Psychology* (Vol. 11). https://doi.org/10.3389/fpsyg.2020.01063
- DeVito, J. a. (2008). The Interpersonal Communication Book. *PsycCRITIQUES*, *32*.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*. https://doi.org/10.1146/annurev.psych.53.100901.135153
- Følstad, A., & Brandtzaeg, P. B. (2018). Chatbots: Changing User Needs and Motivations. *SINTEF, Norway*, *25*(5).
- Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11551 LNCS*. https://doi.org/10.1007/978-3-030-17705-8_13

Annexure

Additional Information | Plagiarism score

- Got a Plagiarism score of 10% when checked with "Turnitin"

REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Thank You!