E-Commerce Based Chatbot System Using Text Mining And Machine Learning Algorithm

Chinnadurai S[1], Pugazheswaran D[2], Ranjith Kumar E[3], Sakthivel V[4], Sivakumar S[5]
[1, 2, 3, 4, 5] Dhanalakshmi Srinivasan Engineering College,Perambalur,AnnaUniversity,India.

Abstract- *In today's e-commerce environment, internet shopping is rapidly expanding. As a result, product recommendation systems have an opportunity to improve. Because users require a relationship with the system. When a relationship develops, the user receives individualized attention and attraction. The system not only tracks and analyses buyer behavior, but it also entices them to return and spend more. The recommendation system eliminates the tedious job of users searching for what they want in an unending category. Instead, they leverage the dialogue to filter out irrelevant information and deliver the product to the customer. While online purchasing offers numerous advantages, it also has limitations and drawbacks that must be considered. When the product purchased and the request made by the customer do not always match, the customer may be disappointed. As consumers' needs change on a daily basis, improving the present functionality of these systems has become a critical aspect. Based on the history of online buying, recommendation systems will be in high demand in the near future. Research is launching a conversation bot that offers products to clients based on their needs. The chatbot essentially takes orders with minimal user involvement and recommends the best product. This can be done on a wide scale, but in this case, the product database is used. Through the chatbot, the customer provides information regarding the perfume. It will also recommend related products based on the user's description.*

Keywords- Chatbot, Recommendation, e-Commerce, Shopping, Product, Purchasing, Big data, Cost, Data Mining, Hadoop, Machine Learning.

I. INTRODUCTION

Big data is high-volume, high-speed and/or high-variety data assets that demand cost-efficient, innovative varieties of informatics that change increased insight, deciding, and method automation. Big data refers to huge complicated structured and unstructured information sets that are quickly generated and transmitted from a large type of sources.

These attributes compose the three Vs of big data:

1. Volume: The large amounts of knowledge being hold on.
2.Velocity: The lightning speed at that information streams should be processed and analyzed.
3. Variety: The various sources and forms from that information is collected, like numbers, text, video, images, audio and text.

Big data is actually the haggle of the three Vs to realize insights and make predictions, thus it's helpful to require a more in-depth cross-check every attribute.

Volume: Massive information is big. Whereas ancient information is measured in acquainted sizes like megabytes, gigabytes and terabytes, big data is hold on in petabytes and zettabytes. To understand the enormity of distinction in scale, take into account this comparison from the Berkeley school of Information: one gigabyte is that the equivalent of a seven-minute video in HD, whereas one zettabyte is capable 250 billion DVDs. This can be simply the tip of the iceberg. In line with a report by EMC, the digital universe is doubling in size each 2 years and by 2020 is expected to succeed in 44 trillion zettabytes. Big data provides the architecture handling this sort of knowledge. While not the acceptable solutions for storing and process, it'd be not possible to mine for insights.

Velocity: From the speed at that it's created to the number of your time required to analyze it, everything regarding big data is quick. Some have represented it as making an attempt to drink from a hearth hose. Corporations and organizations should have the capabilities to harness this information and generate insights from it in time period, otherwise it isn't terribly helpful. Data processing permits call manufacturers to act quickly, giving them a leg up on the competition. Whereas some varieties of information can be batched processed and stay relevant over time, abundant of huge information is streaming into organizations at a clip and needs immediate action for the most effective outcomes. Sensing element information from health devices could be a nice example. The power to instantly method health information will give users and physicians with probably life-saving data.

Variety: Roughly 95% of all big information is unstructured, that means it doesn't match easily into a simple, ancient model. Everything from emails and videos to scientific and

meteorologic information will represent an enormous information stream, each with their own distinctive attributes. This article guides a stepwise walkthrough by Experts for writing a successful journal or a research paper starting from inception of ideas till their publications. Research papers are highly recognized in scholar fraternity and form a core part of PhD curriculum. Research scholars publish their research work in leading journals to complete their grades.In addition, the published research work also provides a big weight-age to get admissions in reputed varsity. Now, here we enlist the proven steps to publish the research paper in a journal.

Uses: The diversity of big data makes it inherently complicated, leading to the need for systems capable of process its numerous structural and linguistics differences. Big data needs specialized NoSQL databases that may store the data in an exceeding approach that does not need strict adherence to a specific model. This provides the flexibleness required to cohesively analyze apparently disparate sources of data to realize a holistic read of what's happening, how to act and once to act. Once aggregating, process and analyzing big data, it is often classified as either operational or analytical information and hold on consequently. Operational systems serve giant batches of knowledge across multiple servers and includes such input as inventory, client information and purchases — the day-after-day information inside a company

II. IDENTIFY, RESEARCH AND COLLECTIDEA

Existing System

Consumers use the analysis or opinion of people as a vital information source. People prefer to get recommendations after they understand a risk in creating a purchase deal call, or after they need to change their buying decision.

Recommendation becomes even additional necessary with in the Internet-based shopping atmosphere wherever customers don't build physical contact with products and face higher cognitive risk. Additionally, e-commerce sites supply a very sizable amount of alternatives since they are doing not have any physical constraint on inventory or shelf area. Hence, customers is also confused by the amount of choices. If the consumer isn't acquainted with the net, the matter becomes even additional serious to solve these issues, many ecommerce sites as using recommender systems to assist their customers make their purchase choices additional with efficiency. A recommender system is associate electronic agent that helps customers to search out the foremost valuable products/services based on their historical preferences or tastes. In existing system implement the collaborative filtering algorithmic rule to ascertain user details and text primarily based question system applied to suggest the product.

Disadvantages

- Time complexity is high.
- Computational process are high.
- Only support text primarily based search,
- Implement machine learning algorithmic rule to classify the text question and recommend the product.

Proposed Systems

Chatbots can bring innovation in on-line help and communication with customers. Due to the expansion of e-commerce, fashion brands are adopting chatbots to supply personal shopper experiences. Analysis in the area of chatbots for e-commerce has addressed technological advancements and shopper behavior, however very little has been done on analyzing chatbot options through a holistic purpose of read. The aim of this project is to supply associate degree interdisciplinary review through a comprehensive categorization of recent studies on the theme and inform future analysis within the space.

During this project we tend to can implement the framework to suggest the product in terms of human computer interaction system which has the natural language process and machine learning algorithmic program with improved accuracy rate. NLP combines computational linguistics that's the rule-based modelling of the human spoken language with intelligent algorithms like applied math, machine, and deep learning algorithms. These technologies along produce the sensible voice assistants and chatbots that you simply could also be utilized in daily life. By artificial means intelligent chatbots, because the name suggests, as created to mimic human-like traits and responses.

NLP or natural language process is massively accountable for enabling such chatbots to know the dialects and undertones of human conversation. IP combined with computing creates a really intelligent chatbot which will answer nuanced queries and learn from each interaction to make better-suited responses consecutive time. The AI chatbots have been developed to help human users on completely different product.

Advantages

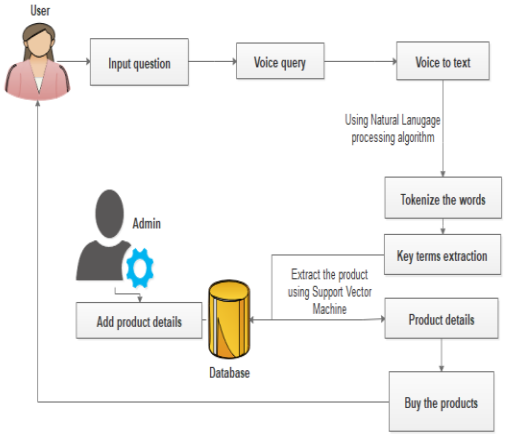
- Easily get the product.
- Build a personality's machine interaction system.

- Extracting information from user typewritten queries.
- The proposed technique provides a value economical and effective thanks to mine knowledge from crowd sourced question respondent datasets.
- Recommending expected results from trained datasets.

III. WRITE DOWN YOUR STUDIES AND FINDINGS

System Architecture

A system architecture is that the abstract model that defines the structure, behavior, and a lot of views of a system. An design description may be a formal description and illustration of a system, organized during a method that supports reasoning concerning the structures and behaviors of the system.



System Overall Architecture

Modules

- Interface Creation
- Post Queries
- Keyword Extraction
- Top-K Results
- Buy the Products

Modules Descriptions

Interface Creations

E-commerce platforms are framed mistreatment recommendation system. This system attempt to recognize customer's behavior and so advocate the products according to their interest. E-commerce framework is employed to shop for the products in on-line to simple retrieval the mobile products. This module is employed to make android and computing device for recommending the best mobiles in specific space. Admin is the responsibility for maintaining the all details is server and server will be design in server. There are two accounts like admin and user account. Admin will log in to the system and post item details. User will login to the system to look at the interface. Then read the merchandise with specified filter. This module is employed to make computing device obtain or post merchandise for users.

Post Queries

A Chatbot or chatter larva could be a code application wont to conduct a web chat voice communication via text or text-to-speech, in office of providing direct contact with a live human agent. A Chatbot could be a style of code which will help customers by automating conversations and move with them through messaging platforms. A talk larva can facilitate the user to induce the input within the kind of search question and so offer the output as a counseled multiple product what user is trying to find. Initial work for analysis is assembling knowledge. Data required for this research contained the details in the form of name, brand, text descriptions, reviews, and a list of notes.

Keyword Extraction

As we've an inclination to square measure exploitation language method, the text data ought to be preprocessed. It covers some tasks like making text data to printed symbol, removing stop words, tokenization, stemming, etc. Lower casing – Lower casing is that the primary step in data pre-processing. The step is easy but necessary. Making the entire text to minuscule is crucial step to urge the consistent output. Tokenization–Once lower casing is finished, tokenization will ensue. It implies that the sentences are divided into substrings known as Tokens. These tokens are employed to understand the words in sentences. Tokenization primarily refers to splitting up a much bigger body of text into smaller lines, words. The various tokenization functions intrinsic into the NLTK module itself. An easy regular expression based tokenize Regexp Tokenizer provided by NLTK was used that splits the text in to punctuations and white spaces words get removed we have a tendency to square measure ready to concentrate on necessary, meaningful words. A tongue toolkit(NLTK) was accustomed load the stop words and deduct them.

Top-K Results

In this module implement Support vector machine rule to supply the merchandise. In machine learning, support vector machines (SVMs) unit supervised learning models with associated knowledge algorithms that analyze data and acknowledge patterns, used for classification and statistical

method. SVMs work supported the principle of Structural Risk diminution Principle. SVMs work okay with text data and chatbot's thanks to the high dimensional input house due to sizable quantity of text choices, linearly divisible data and additionally the prominence of skinny matrix. It's one among the foremost popularly used algorithms for text classification and intent identification. This permits U.S. to assess but ostensibly associate input is to being in one category or another. Cross validation is that the foremost typical manner of testing this rule, assessing but sharp the output created from this manner area unit supported coaching and check sets. Exactitude and revised metrics are wont to assess the process of this model.

Buy the Product

Recommendation systems have return an extended manner since cooperative filtering and content based mostly systems. Considering comments as a vital piece of date and time, that must be processed to separate info outofit and possibly mix its use cases with alternative recommendation systems. Since user comments area unit an immediate kind of knowledge from the user it contains vital keywords supported that more recommendations are often created. This algorithmis typically enforced not solely to AN Ecommerce application however conjointly various completely different applications of constant kind wherever users and a service or product is concerned. When the results, user obtain the product and updated into databases.

System Testing

Testing may be a set activity that may be planned and conducted consistently. Testing begins at the module level and work towards the combination of entire computers based mostly system. Nothing is complete while not testing, because it is significant success of the system.

Testing Objectives:

There area unit many rules that may function testing objectives, they are,

1. Testing may be a method of execution a program with the intent of finding an error.
2. A decent action is one that has high chance of finding associate undiscovered error.
3. A no-hit take a look at is one that uncovers associate undiscovered error. If testing is conducted with success in step with the objectives as declared above, it might uncover errors within the code. Additionally testing demonstrates that software functions seem to the operating in step with the specification, that performance needs seem to possess been met.

There area unit 3 ways to check a program

1.For Correctness.
2.For Implementation potency.
3. For machine quality.

Tests for correctness square measure alleged to verify that a program will precisely what it absolutely was designed to try to. This is often way more troublesome than it's going to initially appear, particularly for giant programs. Tests for implementation potency arrange to notice ways in which to form a correct program quicker or use less storage. It's a code-refining method, which reexamines the implementation section of rule development. Tests for computational complexness quantity to AN experimental analysis of the complexness of AN rule or AN experimental comparison of two or a lot of algorithms, which solve identical downside. The data is entered all told forms singly and whenever a slip-up occurred, it's corrected instantly. A high quality team deputed by the management verified all the mandatory documents and tested the package where as entering the info in the slightest degree levels. The event method involves numerous varieties of testing. Every take a look at kind addresses a selected testing demand. The most common forms of testing concerned within the development method are:

1. Unit Testing
2. System Testing
3. Integration Testing
4. Functional Testing

IV. CONCLUSION

Recent years have shown the rise of chatbots that are in parallel with the advancement of AI (AI). fashionable chatbot vogue can generally be classified into several base elements that are:1) Data(open or shut domain), 2) Response generation (retrieval or generative), 3)Text processing (vector embedding), and 4)Machine learning(ML)model (usually victimization neural network). throughout this project, we have got presented a novel implementation of a product recommendation system supported chatbots. The reason for building such normal system is to make the system accessible to more platforms. This project presents Associate in Nursing discipline vogue and implementation review of ve fashionable chatbot systems. The aim for this review is to supply associate degree outline sense with relevance vogue follow to boot as implementation strategy in fashionable Chat-bot systems. This gift natural language method engine trains its classifier from

the classified work information provided by the admins. Also, it's supported SVM.

V. FUTURE ENHANCEMENT

In the future, we'll extend the framework to implement in various applications and recommend the merchandise supported multiple choices like likes, user preferences so on.

REFERENCES

[1] Anand, Deepa, and Kamal K. Bharadwaj. "Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities." Expert systems with applications 38.5 (2011): 5101-5109.

[2] Angappan, A. Kumaresan, et al. AI Based Healthcare Chatbot System. No. 5260. EasyChair, 2021.

[3] Arooba Khalid Bashir Shaik, Farah and Heba Mahmoud Ismail. "Covisstance chatbot." The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research. 2021.

[4] Aqib Anwar, Jameel, Hashim Khan and Umar. "Doctor Recommendation Chatbot: A research study: Doctor Recommendation chatbot." Journal of Applied Artificial Intelligence 2.1 (2021): 1-8.

[5] Ayanouz,. "A smart chatbot architecture based NLP and machine learning for health care assistance." Proceedings of the 3rd International Conference on Networking, Information Systems & Security. 2020.

[6] Bobadilla, Jesús, et al. "Recommender systems survey." Knowledge-based systems 46 (2013): 109-132.

[7] Bobadilla, JesúS, et al. "A collaborative filtering approach to mitigate the new user cold start problem." Knowledge-based systems 26 (2012): 225-238.

[8] ByeongJin, et al,Ye. "Development of a Chatbot Program for Follow-Up Management of Workers' (2021): 2170.

[9] Chenghua, Lin, et al. "Weakly supervised joint sentiment-topic detection from text." IEEE Transactions on Knowledge and Data engineering 24.6 (2011): 1134-1145.

[10] Christos K. Georgiadis, Nikolaos and Polatidis. "Recommender systems: The Importance of personalization in E-business environments." International Journal of E-Entrepreneurship and Innovation (IJEEI) 4.4 (2013): 32-46.

[11] Gang ,Wang, et al. "Sentiment classification: The contribution of ensemble learning." Decision support systems 57 (2014): 77-93.

[12] J.Jinu, et al, Sophia. "A survey on chatbot implementation in health care using NLTK." Int. J. Comput. Sci. Mob. Comput 9 (2020).

[13] Jitgosol, PanidaBoonchai, Sumonta Kasemvilas and Yuvadee. "Designing an HR chatbot to support human resource management." (2019): 165.

[14] Karthik, Pallavi Shetty and Saligrama. "AI Enabled Healthcare Chatbot Systems for Hospital Web Applications" 2020.

[15] Kiyoaki Shirai, Nguyen and Thien Hai. "Topic modeling based sentiment analysis on social media for stock market prediction." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015.