AI CHAT BOT FOR eCOMMERCE

Introduction **1**

Challenges 2

Proposed Solution 2

Brief Conclusion 2

Business Understanding: **3**

Problem Statement: **3**

Main objective. 3

Specific objectives 4

Business and Data Understanding **4**

Data Source 4

Data Description. 4

Data Analysis **5**

Modeling **5**

Evaluation **5**

Recommendations **5**

Introduction

Artificial intelligence (AI) chatbots are apps or interfaces that can carry on human-like conversation using natural language understanding (NLU) or natural language processing (NLP) and machine learning (ML). They use large language models (LLMs) to generate responses to text. This can be used as a virtual response assistant.

In the ever-evolving landscape of artificial intelligence, the integration of AI chatbots has emerged as a transformative solution with wide-ranging applications. This project centers around the development of versatile AI chatbots that serve as a responders. By harnessing the power of natural language processing and machine learning, we aim to create intelligent conversational agents capable of enhancing user experiences in diverse domains.

Project overview

The AI Chatbot for Ecommerce Capstone project is designed to revolutionize client-business interactions by introducing a sophisticated Frequently Asked Questions (FAQ) system. Our primary goal is to offer seamless and efficient assistance to both clients and business owners through the implementation of advanced machine learning algorithms, including Naive Bayes and the Random Forest Classifier. To augment the chatbot's capabilities, we plan to integrate pretrained models that will ensure a comprehensive understanding of user queries, enabling accurate and timely responses. The project will be executed in several phases, starting with rigorous data preprocessing to ensure data quality and consistency. We will then proceed to identify and extract relevant features from integrated ecommerce data to enhance model training. Utilizing historical data with known client-business interactions, we will train machine learning models, with a focus on optimizing the system for scalability to accommodate a growing user base. Emphasis will be placed on providing transparent and understandable model outputs, facilitating effective decision-making.Our

success criteria include measuring the chatbot's accuracy and responsiveness, evaluating scalability.

Challenges

eCommerce’s popularity has increased in recent years. It’s now fast becoming the go-to shopping platform for many people. A report by Statista reveals that its global sales are projected to reach a whopping $8.1 trillion by 2026. These staggering numbers show eCommerce’s great potential. As your eCommerce business grows, you’ll receive more customer inquiries via email, phone, and social media. It can be challenging to manage these inquiries efficiently, especially if you’re a small business with limited resources. For example, a customer who purchased a product from an eCommerce website has questions about its specifications. He sends an email to the customer service team asking for clarification. However, the inquiry was not responded to on time due to mismanagement, as a result, the client felt neglected and unheard, leading the customer to perceive the brand negatively.

Proposed Solution

Investing in a customer service chatbot that seamlessly integrates all communication channels is a strategic move for any ecommerce store. By adopting such technology, you streamline your customer support operations, offering a centralized solution for efficiently managing client inquiries across various channels. This not only enhances the customer experience but also boosts team productivity. With features like canned responses and templates, customer enquiries can swiftly address common questions, saving valuable time and ensuring consistent responses. Additionally, the platform, ensures timely resolution and improved customer satisfaction. In essence, embracing a chatbot-integrated customer service platform empowers your ecommerce store to deliver exceptional support, build customer trust, and ultimately drive business growth.

Brief Conclusion

In conclusion, implementing a chatbot for customer support offers a transformative solution for businesses, streamlining communication processes and enhancing

customer satisfaction. By leveraging artificial intelligence and natural language processing capabilities, chatbots provide immediate assistance to customers, addressing their inquiries and resolving issues efficiently around the clock. This not only reduces response times but also enables businesses to handle a high volume of queries simultaneously, scaling support operations effectively. Moreover, chatbots can offer personalized recommendations, gather feedback, and analyze customer interactions, allowing businesses to continuously improve their services. Overall, integrating chatbots into customer support strategies empowers businesses to deliver seamless, proactive, and tailored assistance, fostering stronger customer relationships and driving long-term success.

Business Understanding:

In the dynamic realm of artificial intelligence, the utilization of AI chatbots has become increasingly integral, presenting numerous opportunities for transformative applications. This project focuses on the development of versatile AI chatbots designed to serve as customer support. Leveraging advanced natural language processing and machine learning techniques, the goal is to craft intelligent conversational agents capable of elevating user experiences across a spectrum of domains.

Problem Statement:

In today's rapidly advancing digital landscape, the integration of chatbots has become a pervasive trend, offering enhanced efficiency and convenience. In the dynamic world, chatbots help do the day-to-day running of customer support.The current landscape of chatbots often falls short in adaptability, lacking the versatility required for diverse roles. The challenge at hand is to address this limitation by creating an AI chatbot framework that can effortlessly resolve customer concerns, customer queries and have meaningful interactions with the end user. Time is a precious commodity and having enough in a day to do the things that matter is important. Time to answer all business messages can be overwhelming, leading to a disconnect between the business and the customer . As a result, they may feel unheard and miss out on

valuable interactions. Thus,enhancing the engagement experience for customers, through AI-powered chatbot designed to provide responsive, context-aware assistance without the limitations of human-operated systems.

Main objective.

To engineer a dynamic and user-centric AI chatbot system for businesses and corporate intranets. This system should not only grasp context effectively but also adapt to user preferences, providing a comprehensive solution for varied scenarios. By pushing the boundaries of virtual interactions, the project aims to redefine user experiences, offering innovative AI-driven chatbot functionalities.

Specific objectives

1.To enhance customer engagement.
2.To increase operational efficiency by avoiding bottlenecks when it comes to response time (Wu et al., 2020).
3. Sales Support in ecommerce by providing product information to potential customers.

Business and Data Understanding

Data Source

Amazon question/answer data **https://cseweb.ucsd.edu/~jmcauley/datasets/amazon/qa/**

Reference

Wu, Q., Ma, J., & Wu, Z. (2020, April). Consumer-driven e-commerce: A study on C2B applications. In *2020 International Conference on E-Commerce and Internet Technology (ECIT)* (pp. 50-53). IEEE. DOI: 10.1109/ECIT50008.2020.00019

Data Description.

This dataset is stored in the "QA_Beauty.json.gz". It contains Question and Answer data comprising of 32936 rows and 6 columns.Here is a brief description of each column:

1. **asin**: This column represents the ID of the product, serving as a unique identifier, such as "B000050B6Z".

2. **questionType**: It denotes the type of question and can take on values like 'yes/no' or 'open-ended', indicating the nature of the inquiry.

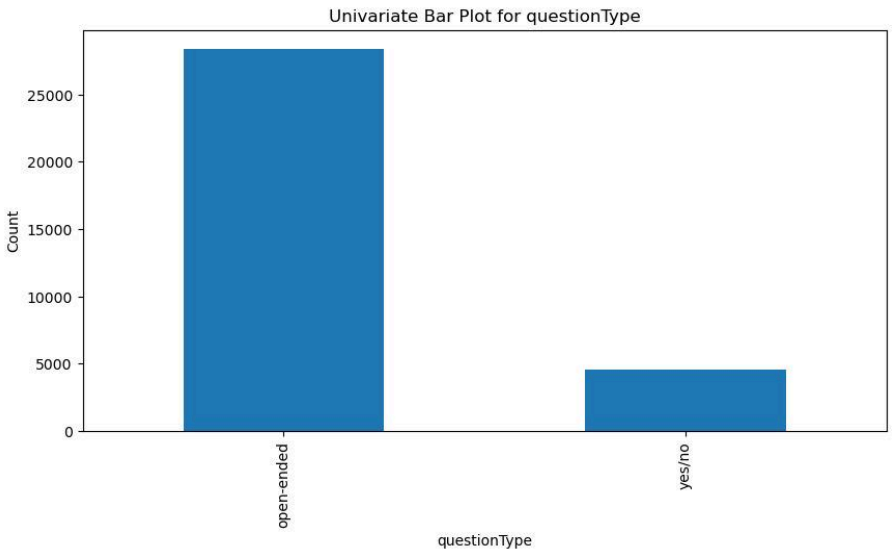
3. askerID: This column contains the ID associated with the individual asking the question, providing a reference to the person initiating the inquiry.

4. **questionText**: It holds the text of the question, providing insight into the specific queries users have about the products.

5. **answers**: This column encapsulates the responses to the questions posed. It could contain a variety of information, including yes/no answers or more detailed responses for open-ended questions.

6. **questionTime**: This column contains timestamp information representing the date and time when each question was asked.

Data Analysis



The graph presented is a univariate bar plot that displays the distribution of question types within a dataset. This dataset appears to be related to an e-commerce setting question/answer data.

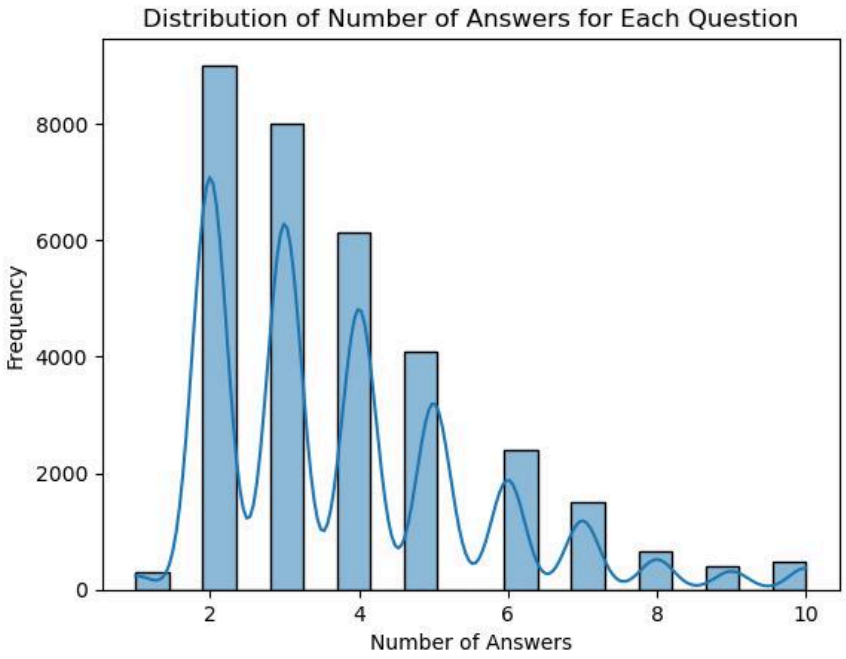
From the bar plot, we can observe two categories of question types:

1. Open-ended: This category has the largest number of occurrences, indicating that it is the most common type of question in the dataset. The bar representing open-ended questions is significantly higher than the one for yes/no questions, suggesting that users prefer asking questions that require a more detailed response rather than just a simple affirmative or negative. This could mean that customers are looking f~~or more in-depth information about products or have queries that cannot be answered with a simple yes or no.

2. Yes/no: This category has a much lower count in comparison to open-ended questions, indicating that such questions are less frequent in the dataset. Yes/no questions are likely to be those that require a straightforward answer without the need for elaboration.

The 'open-ended' question type exhibits the highest number of occurrences in the dataset, suggesting that users frequently engage in inquiries that prompt detailed and unrestricted responses. This prevalence highlights the users' inclination towards seeking comprehensive information or explanations.

In a customer service chatbot context, this data could be used to prioritize the development of features that support open-ended questions, ensuring that the bot is capable of handling detailed and complex user queries effectively.



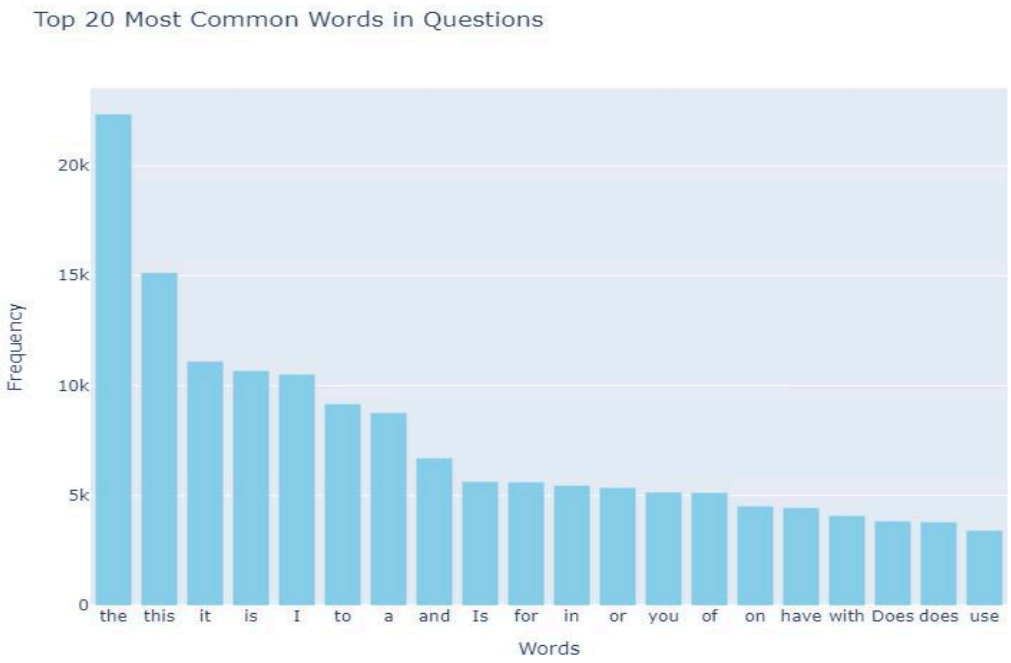
The histogram shows that the majority of questions receive a specific number of answers, with the number two being the most common, as indicated by the highest bar on the graph.

With the added context in the caption, it suggests that the dataset contains a significant number of questions that receive exactly two answers, pointing towards a pattern of binary or dual-response format. This could imply that many users are satisfied with two answers to their questions, or that the system or process that generates these answers tends to provide two by default or as a limit.

The prevalence of two answers could be due to several reasons, such as:

- Users often pose questions that can be sufficiently answered with a couple of responses, or they tend to prefer a second opinion.
- The platform may encourage or limit users to provide two answers, which could be a design choice to prevent information overload or to simplify the decision-making process for readers.
- The nature of the questions could be such that they naturally elicit a binary response, representing a comparison or a choice between two options.

Understanding this pattern is crucial for the design and implementation of an AI chatbot for customer service in e-commerce.



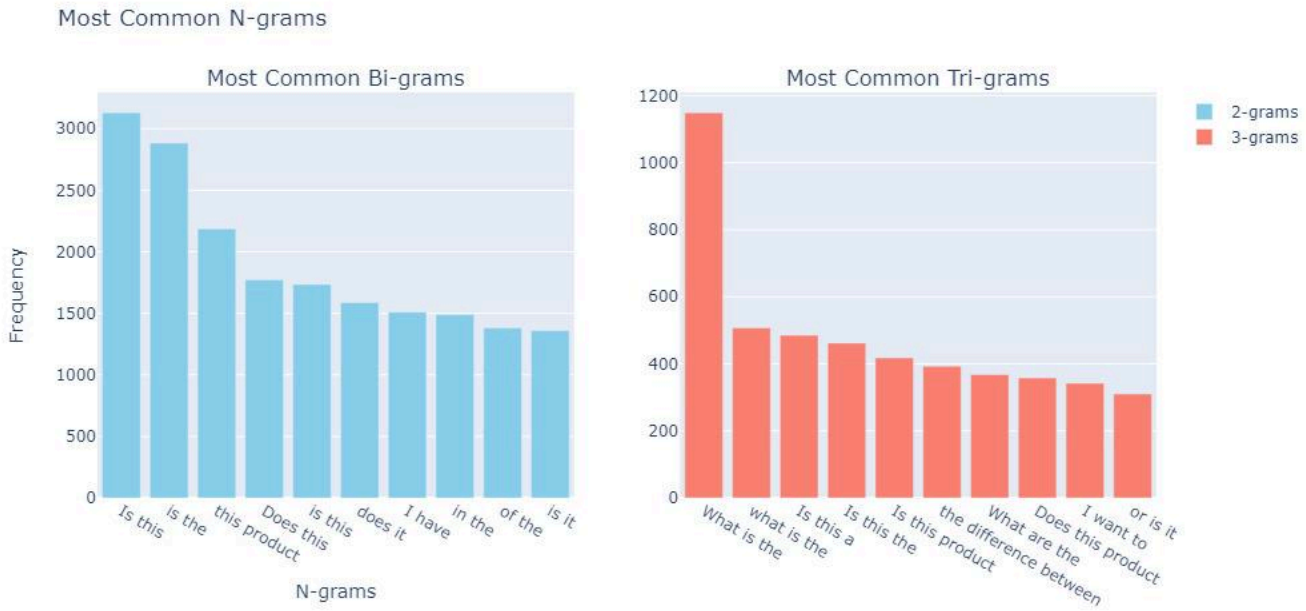
The visualization displays the top 20 most frequently occurring words found in questions from the dataset. These types of words are typically used in text analysis to understand the common language or topics that are being discussed.

Here are some key observations:

- "the" and "this" are the most common words, with "the" being the most frequent. These are articles in English and are commonly used in many types of sentences, which explains their prevalence.
- The list is dominated by function words, such as articles ("the", "a"), pronouns ("this", "it"), conjunctions ("and"), prepositions ("to", "for", "of", "on"), and auxiliary verbs ("is", "does"). These words are often used to form questions and are typically frequent in most English language corpora.
- There are also some content words like "use", which may be relevant to the e-commerce context of the dataset, potentially indicating that questions about how to use products are common.

- The presence of words like "have", "does", and "use" towards the end of the list suggests that questions in the dataset may involve ownership ("have"), inquiries about functions or features ("does... have"), and usage ("use").

This chart is useful for understanding the structure of questions that customers are asking.



Most Common Bi-grams: The left chart shows the frequency of bi-grams:

- "is this" is the most frequent bi-gram, suggesting that many questions start with an inquiry directly about a product or feature.
- Following "is this", the next most common bi-grams include "is the", "does this", "have it", "in the", and "of the". These bi-grams reflect common question patterns, where users are asking about specific attributes of a product, its availability, or seeking to understand its placement within a certain context.
- The least common bi-gram among the top shown is "is it", which could be part of questions about product confirmation or verification.

Most Common Tri-grams: The right chart displays the frequency of tri-grams:

- "what is the" is the most frequent tri-gram, which is often a lead into a detailed question about a product or service.
- Following this, the other common tri-grams include "is this the", "is the this", and "what is this". These suggest that users are looking for specific information or verification about products.

- The least common tri-gram shown is "or is it", which might be used in comparative or alternative inquiries about products or features.

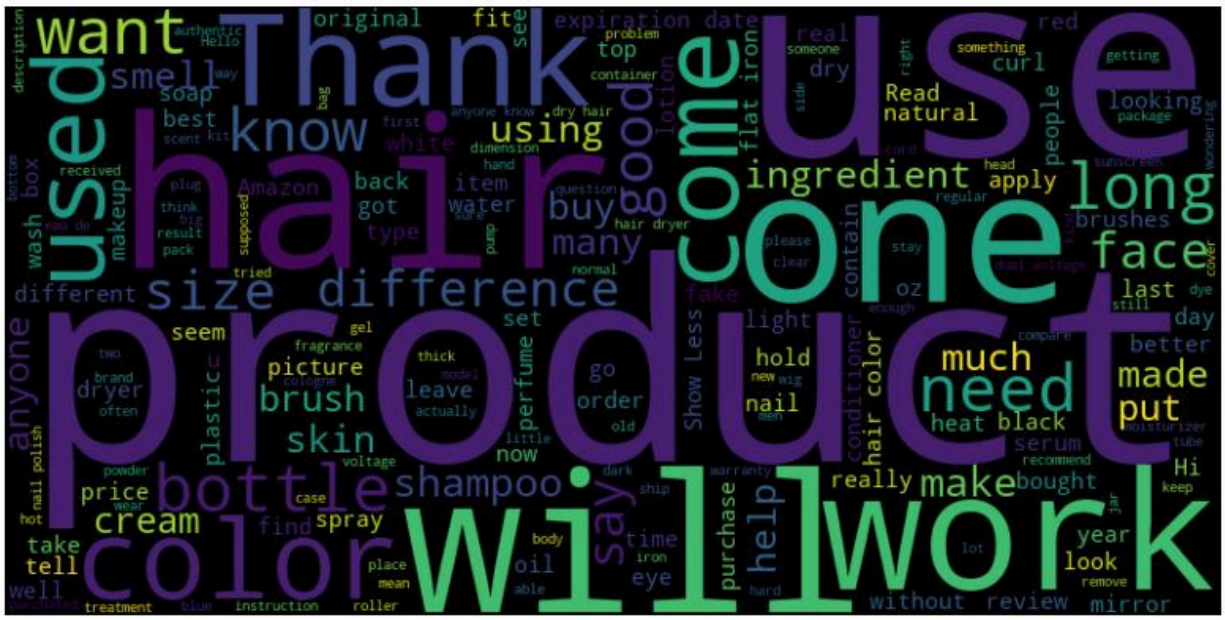
Notable Words: The caption also mentions notable words such as "product", "use", "work", "will", "Thanks", and "One". These content-specific words can provide insights into common themes in customer queries, such as questions about how to use a product ("use"), its functionality ("work"), future availability or action ("will"), gratitude or follow-up ("Thanks"), or selecting an option ("One").

Implications for AI Chatbot Development: This n-gram analysis is very useful for designing an AI chatbot for e-commerce customer service:

- Understanding common bi-grams and tri-grams can help in training the chatbot to recognize frequent patterns in customer inquiries.
- The chatbot can be programmed to trigger certain responses based on these common n-grams, enhancing its ability to address the most frequent types of questions.
- Content words like "product", "use", "work", "will", "Thanks", and "One" might be critical in understanding the intent behind a customer's question and providing accurate responses.

Thus, the n-gram analysis helps in understanding the structure and subject of customer queries, which is crucial for building an effective AI chatbot that can provide relevant and helpful answers.

Word Cloud for Question Text



In this word cloud, we can identify several prominent words:

- "**product**": Its large size suggests that it is one of the most frequently mentioned words, indicating that many questions are directly related to specific products.
- "**use**": Also quite large, implying that questions about how to use items or the uses of different products are common.
- "**work**": Likely related to questions about the functionality or effectiveness of products.
- "**will**": This could be part of questions about the future effects of using a product or the product's longevity.
- "**Thanks**": This appears to be a common sign-off in questions, showing politeness from the askers.
- "**One**": Might be used in the context of selecting a product or choosing between options.

Other noticeable words include "size", "color", "ingredient", "buy", "face", "skin", "brush", "cream", "shampoo", "makeup", "long", "need", and "help". These words suggest that the questions are related to personal care products and that customers are concerned with product attributes (like size and color), ingredients, purchase decisions, and usage instructions.

From an AI chatbot development perspective, the word cloud informs the areas where the chatbot should have robust knowledge. For instance, the chatbot should be able to answer questions about product specifications ("size", "color"), ingredients, how to use products ("use", "apply"), and possibly the efficacy ("work") of products. It also suggests that customers appreciate courteous interaction ("Thanks"), which should be part of the chatbot's response templates.

Modeling

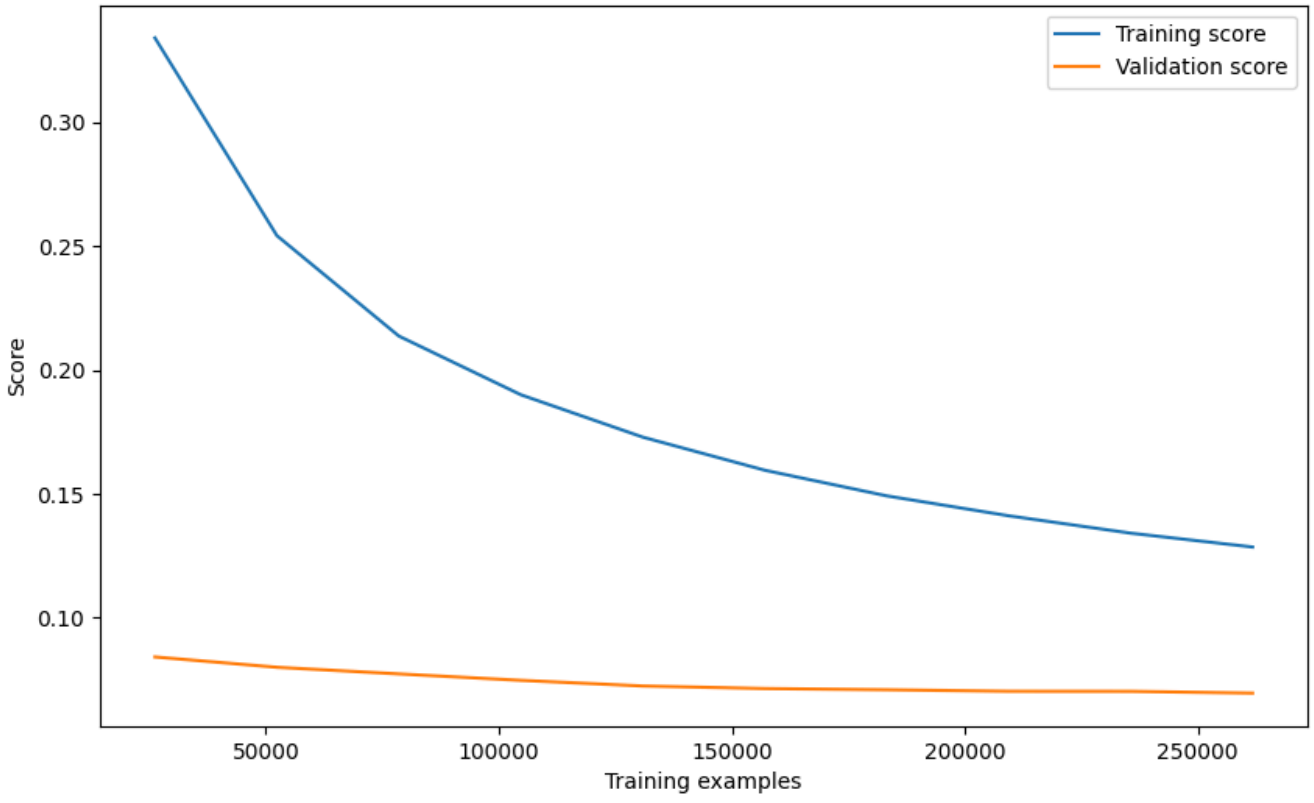
Multinomial Naive Bayes Model

Accuracy: 0.06140983486328244

The MultinomialNB model makes an independence assumption and suits discrete data.

Evaluation

The low accuracy indicates it did not generalize well. A look into the classification report confirms the low scores across all the intent classes. The macro average precision is 0.05, recall is 0.06 and F1 is 0.06, highlighting that the model struggled to correctly classify examples from the minority classes. Most classes have precision and recall scores in the 0.02 to 0.07 range, with just the "Personal Care" intent slightly higher at 0.12 precision and 0.20 recall. Such large class imbalances showcase the model's inability to learn effectively from skewed datasets. The model's average cross-validation accuracy over 5 folds is 0.07, which aligns with the low test accuracy. This confirms the model's subpar performance irrespective of train-test split.



Random Forest

Random Forest Test Accuracy: 0.8800475059382423

In Random Forest classifier is trained and evaluated after processing the data, with the model focusing on the features 'questionText' and 'answerText' to predict the tag 'questionType'. The dataset is split into training and testing sets, and the classifier is created with 100 trees. The accuracy is then calculated using the test set, resulting in a high accuracy of 88.004%. This indicates that the model, trained specifically on the textual content of questions and answers, performs well in classifying instances into different question types. It's important to note that the success of the model relies on the features chosen for training, and the achieved accuracy provides insight into the effectiveness of the classifier in this particular context. Nonetheless, considering other evaluation metrics and potential hyperparameter tuning might further optimize the model for the specific characteristics of the classification problem.

Sequential Neural Network Model

Neural Network Test Accuracy: 0.8640924096107483

Sequential neural network model is constructed using the TensorFlow Keras library for a 'questionType' classification task based on textual data from 'questionText' and 'answerText.' The text is preprocessed by tokenizing and lemmatizing using the NLTK library. The lemmatized text is then converted into bags of words, and the output rows are one-hot encoded. The Count Vectorization technique is employed to convert the text data into numerical format, utilizing a reduced vocabulary size. The resulting numerical data is split into features (X) and labels (Y), followed by a further division into training and testing sets. The neural network model is designed with an input layer of 128 neurons, a dropout layer to mitigate overfitting, a hidden layer with 64 neurons, another dropout layer, and an output layer with softmax activation for multiclass classification. The model is compiled with categorical crossentropy loss, the Adam optimizer, and accuracy as the metric. Subsequently, the model is trained over five epochs, with a batch size of 32 and a validation split of 20%. The trained model is then evaluated on the test set, yielding a neural network test accuracy of approximately 86.41%. This accuracy metric underscores the proficiency of the neural network in effectively classifying 'questionType' based on the provided textual features.

Recommendations

1. **Optimize for Scalability**: As the eCommerce platform grows, the volume of customer inquiries will increase. It's vital to ensure that the chatbot system is scalable, capable of handling a large number of queries without compromising response time or accuracy.
2. **Personalization:** Implement features that allow the chatbot to offer personalized

recommendations and responses based on the user's browsing and purchase history. This can enhance the customer experience and potentially increase sales conversions.

3. **Feedback Loop Integration**: Establish a mechanism for collecting user feedback on chatbot interactions. This data can be invaluable for continuous improvement of the chatbot's responses and functionalities.
4. **Multilingual Support:** Considering the global reach of eCommerce, adding multilingual support to the chatbot can make it accessible to a wider audience, thereby enhancing customer support and inclusivity.

Github Repository: https://github.com/MwangiWambugu/AI-chatbot-for-eCommerce-Store