Automated Content Moderation in a Brazilian Marketplace

Ana Claudia Zandavalle
Americanas S.A.
Florianópilis, Santa Catarina, Brasil
ana.zandavalle@americanas.io

Victor do Nascimento
Americanas S.A.
Santos, Brasil
victor.rnascimento@americanas.io

Carolina Gadelha
Americanas S.A.
Santos, Brasil

Tatiana Gama
Americanas S.A.
Rio de Janeiro, Rio de Janeiro, Brasil
tatiana.gama@americanas.io

Fernando Zagatti
Americanas S.A.
São Carlos, Brasil
fernando.zagatti@americanas.io

Lucas Nildaimon
Americanas S.A.
São Carlos, Brasil
lucas.nildaimon@americanas.io

João Gabriel Melo Barbirato
Americanas S.A.
São Carlos, São Paulo, Brasil
joao.barbirato@americanas.io

Livy Real
Americanas S.A.
Mongaguá, São Paulo, Brasil
livy.coelho@americanas.io

Abstract

Clarifying doubts can become decisive when shopping on e-commerces platforms. Considering the relevance of user generated content, this work aimed to develop an internal hybrid system, composed of machine learning models alongside a rule-based module, to moderate customers' questions and sellers' answers in one of the biggest marketplaces in Brazil.

Keywords: content moderation, Portuguese, questions and answers, user generated content, e-commerce, marketplace.

1 Introduction

In e-commerce platforms, user-generated content has been gaining more and more relevance to its business since it increases the consumer confidence level to conclude online purchases [2]. As one of the largest marketplace in Latin America, Americanas Marketplace has its business based on online shopping, connecting sellers and final customers through its platform. One type of content that facilitates this interaction is the question and answer (Q&A) technology: a system that allows customers to publicly ask questions in natural language, in this case, mostly Brazilian Portuguese, in the product page and to publicly receive answers from a seller. Figure 1 illustrates this section on the product page.

The questions are mainly asked in the pre-purchase stage and their content must focus on doubts about product features, especially those that have not been specified in the product description or in the product technical data sheet.

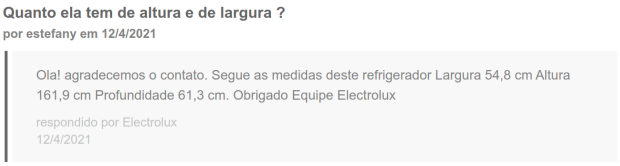

Figure 1. Q&A section.
Source: https://www.americanas.com.br/produto/111957454. Last accessed in jun. 30, 2022.

This kind of content can be really useful not only to the one asking it, but also for other clients that would like to have the same information about this given product. However, it is not unusual for customers and sellers to approach other topics, such as delivery time, freight costs or personal issues in general. Such contents are considered inappropriate to be displayed in the Q&A section of the product page, since they are not relevant to all customers and the answers for them are not stable. Therefore, certain questions and answers should be blocked and not be displayed in the platform.

Due to the large number of interactions that Americanas receives, around one million questions per year, manual moderation, that is to say, to filter what should or should not be available in the product page, becomes impracticable. In that way, scalable solutions are fundamental, since clients need their questions answered as fast as possible to continue their purchases.

Although most research focuses on question-answering (QA) communities, such as `Stackoverflow` [3, 5, 7], there are works related to Q&A, such as how this system affects customers reviews [1]; the impact of information quality on customers' purchase intention [9]; the economic impact of the Q&A section on the product page [6], among others.

In: WebMedia in Practice, Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

ISSN 2596-1683

However, there is a lack of research focusing on Q&A content moderation.

Whereas an online product review represents a unilateral communication channel that allows customers to share post-purchase experiences, the Q&A system permits a bilateral interaction between users that ask questions and sellers that answer them. It enables the reduction of customer uncertainty and enhances purchase intention [9].

Considering the relevance of this topic, this work aimed to automatize Q&A content moderation on e-commerce product pages. To do so, a hybrid system, composed of machine learning models alongside a rule-based module, was created, tested and implemented. The main purpose of this project was to achieve better, cheaper and faster results when compared to the prior moderation method used by Americanas Marketplace. Previously, a third-party company was responsible for this content moderation, so we were also looking to internalize the operation of this system.

This work is divided as the following: section 2 describes the methodology used, alongside the business rules, the hybrid system details and the datasets used to build the models. Next, section 3 presents the results, the evaluation metrics and the model error analysis. Finally, the section 4 shows the conclusions and future research.

2 Methodology

Given a customer's question or a seller's answer, the automatic moderator must define whether or not it should be published on the product page. This work considers the following criteria due to business rules:

- The question should be restricted to the characteristics of the product.
- The question cannot be related to shipping, price, complaints, problems related to the purchase journey.
- The question/answer must not contain bad words.
- The answer cannot head or influence the costumer to purchase the product in a different platform.

In addition, the solution aims to have better metrics in Q&A moderation than the third party company had and to have a prediction time lower than the previous solution.

Considering these directives, this work divides the solution into two parts: a specific moderation for the questions and a second moderation only for the sellers' answers. For each part of the solution it was developed a hybrid model composed by two modules. The first one is a rule-based module composed of a list of prohibited words (`Blocklist`), described in section 2.1; the second one is a module composed of a machine learning model for automatic classification, detailed in section 2.2.

2.1 Blocklist Module

The `Blocklist` module consists of a list of n-grams that are very often used in contexts that always should be blocked, such as: delivery, freight, price, market competitors, inappropriate content and bad words. If the text contains any n-grams in this list, the question, or answer, is automatically blocked. In this case, to the customers, the submit button in the Q&A web interface becomes unavailable and a user feedback is instantly reported showing the Q&A section rules. This list was composed using regular expressions and manual annotation considering business and linguistics knowledge. A comparative analysis was performed between a sample randomly extracted and manual annotated data and results are shown in 3.

2.2 Moderation Module

The Moderation Module consists of a training of two Deep Learning models. One is responsible for moderating the questions sent by customers, while the other is responsible for moderating the responses sent by sellers. Both models have the same architecture, a Bidirectional Encoder Representations from Transformers (BERT) [4] model pre-trained in Brazilian Portuguese [11], followed by a dense layer to conduct the fine-tuning. The output of each model is binary: it rejects or accepts the content to be moderated.

2.2.1 Datasets. The data used to develop the models were manually categorized into questions/answers that can be published on the website (ACCEPTED) and the ones that cannot (REJECTED). Both datasets are composed only by questions/answers that were not blocked by the `Blocklist` module. The `Questions` dataset has 3781 questions, with 2211 labeled as ACCEPTED and 1570 as REJECTED. The `Answers` data is composed of 5238 answers, with 4277 ACCEPTED and 961 REJECTED samples.

Both questions and answers datasets were split into train, validation and test sets keeping the proportion of 64% for training, 16% for validation and 20% for testing. Given the unbalanced data, the data split considered stratified sampling, so all the subsets keep the same proportion of the class of the original sets.

2.2.2 Training. The models were trained until the validation loss stops to decrease for 5 epochs. One step before training the models a random search was conducted to define the best batch size, learning rate and whether to use Binary Cross Entropy Loss with class weights or Focal Loss[8] to deal with the unbalanced classes.

The **Question model** was trained with an Adam optimizer, learning rate of 2.58×10^{-5} and batch size of 8. The Loss Function used was the Cross Entropy Loss. To deal with the unbalanced classes, the weight for the classes was calculated so that each example contributed in an inversely proportional way to the class frequency in the loss calculation.

The **Answers model** was trained with Adam optimizer and a learning rate of 4.30×10^{-5} and batch size of 8. To deal

Table 1. Questions and Answers models metrics for both REJECTED and ACCEPTED classes.

	REJECTED			ACCEPTED		
Model	Precision	Recall	F1- Score	Precision	Recall	F1- Score
Questions (third-party company)	0.912	0.653	0.761	0.887	**0.977**	0.930
Questions (Americanas)	**0.918**	**0.821**	**0.867**	**0.938**	0.974	**0.956**
Answers (third-party company)	0.846	0.083	0.152	0.861	**0.997**	0.924
Answers (Americanas)	**0.949**	**0.561**	**0.705**	**0.928**	0.995	**0.960**

with unbalanced classes, the model was trained with Focal Loss.

3 Results

Regarding the `Blocklist` Module, we reached the following accuracies: 82% and 95% for questions and answers, respectively. This discrepancy in the results is justified because the content of the answers is more homogeneous and the sellers are often more careful with their writing than the normal customer. Based on these results, we kept the ruled based module in our pipeline, since it already satisfactorily solved part of our issue with a low computational cost and in real time.

Regarding the `Moderation Module`, Table 1 shows the achieved metrics.

Our **Question model** and **Answer model** were able to increase the F1-Score in both classes when compared to the third-party company moderation. Therefore, our solution was able to overcome the performance the third-party company and it was ready to be tested in production.

3.1 Model Error Analysis and Testing in Production

In addition to the test set metrics shown above, it was necessary to evaluate the performance of the developed models in the real world scenario. Thus the new versions of the service were started in a shadow deployment [10], which means that requests are sent to both the current and new versions, but only the current stable version delivers responses to the final user. It allows the monitoring of the new service version without affecting customers. After a two-week running test, two samples, one of questions and one of answers, were randomly extracted with the predictions of the four models (Questions Americanas, Questions third-party, Answers Americanas and Answers third-party). To identify the best one for each context, both samples were blindly human-annotated. We found that the `Questions Americanas` model obtained a 30% higher accuracy compared to the third-party company's system, in addition to reaching an average moderation time of 38 milliseconds, against 47 minutes. For the `Answer Americanas` model, there was a higher accuracy of 76% and an average moderation time of 46 milliseconds, against 30 minutes performed by the previous solution. From these results, the two Americanas models developed could be officially the solution in production.

4 Conclusions

We shared how we implemented a new Q&A automated moderation system in Americanas Marketplace. Our service is a hybrid system composed of machine learning models and a rule-based module. The solution considered the business problem and the business rules to achieve the best results in terms of information quality and user experience. The results showed that the Americanas solution performance is better in terms of quality and faster than the third-party company solution used previously. Future works will focus on lowering the dependency of a rule-based module in the system.

References

[1] Shrabastee Banerjee, Chrysanthos Dellarocas, and Georgios Zervas. 2021. Interacting user-generated content technologies: How questions and answers affect consumer reviews. *Journal of Marketing Research* 58, 4 (2021), 742–761.

[2] Yahui Chen, Dongsheng Liu, Yanni Liu, Yiming Zheng, Bing Wang, and Yi Zhou. 2022. Research on user generated content in Q&A system and online comments based on text mining. *Alexandria Engineering Journal* 61, 10 (2022), 7659–7668.

[3] Erik Choi and Chirag Shah. 2016. User motivations for asking questions in online Q&A services. *Journal of the Association for Information Science and Technology* 67, 5 (2016), 1182–1197.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[5] Marcos Menon José, Marcelo Archanjo José, Denis Deratani Mauá, and Fábio Gagliardi Cozman. 2022. Integrating Question Answering and Text-to-SQL in Portuguese. In *Computational Processing of the Portuguese Language*, Vládia Pinheiro, Pablo Gamallo, Raquel Amaro, Carolina Scarton, Fernando Batista, Diego Silva, Catarina Magro, and Hugo Pinto (Eds.). Springer International Publishing, Cham, 278–287.

[6] Warut Khern-am nuai, Hossein Ghasemkhani, and Karthik Kannan. 2017. How questions and answers shape online marketplaces: The Case of Amazon answer. *50th Hawaii International Conference on System Sciences* 50, 1 (2017), 853–862.

[7] Ashish Kulkarni, Kartik Mehta, Shweta Garg, Vidit Bansal, Nikhil Rasiwasia, and Srinivasan Sengamedu. 2019. ProductQnA: Answering User Questions on E-Commerce Product Pages. In *Companion Proceedings of*

The 2019 World Wide Web Conference (San Francisco, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 354–360. https://doi.org/10.1145/3308560.3316597

[8] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2020. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 2 (2020), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

[9] Xiao-ping Liu and Wen-xiang Deng. 2018. The Researches on the Impact of Community Q&A Information Quality on Consumers' Purchase Intention. *Journal of Mathematics and Informatics* 14 (08 2018), 45–52. https://doi.org/10.22457/jmi.v14a6

[10] Alex Serban, Koen van der Blom, Holger Hoos, and Joost Visser. 2020. Adoption and Effects of Software Engineering Best Practices in Machine Learning. In *Proceedings of the 14th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (Bari, Italy) *(ESEM '20)*. Association for Computing Machinery, New York, NY, USA, Article 3, 12 pages. https://doi.org/10.1145/3382494.3410681

[11] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In *Intelligent Systems*, Ricardo Cerri and Ronaldo C. Prati (Eds.). Springer International Publishing, Cham, 403–417.