Improving the KYC Process

Daniel Cristian Dumitrescu

Executive summary

This research looks into the KYC process and proposes some recommendations for the improvement of the total 'pass rate'. The main driver towards the failure of the KYC is the document check. Many customers are uploading low-quality ID pictures, so better instructions on the uploading page should be provided. Another problem identified is that some IDs are expired. This problem can be easily resolved by asking the customer to input the ID's expiration date before uploading the picture of it. If the ID is expired, do not allow the customer to start the KYC process. The potential impact, if both the recommendations are applied, is an increase in the overall 'pass rate' from 75% to 85%.

1. Introduction

As X is regulated by the FCA, it has the responsibility of checking the identity of all new customers who want to open an account. This phase is called the Know Your Customer process and involves two checks. The document check verifies that the document is valid, and the facial similarity check verifies that the face in the picture is the same as the one in the ID. The customer is only able to join X when both checks are passed.

The 'pass rate' is defined as the number of customers who pass both checks divided by the number of customers who attempted the process. Over the last period of months, this has decreased consistently. The goal of this report is to identify the causes and to provide some recommendations for an improvement in the 'pass rate'.

2. Methodology

The available dataset is stored in two CSV files: facial_similarity_reports.csv and doc_reports.csv. For the data pre-processing and analysis steps the Python programming language is being used with the following libraries: Pandas, NumPy, Matplotlib, JSON, Seaborn. For the visualisation, bar plots, heat maps, time-series and correlation matrixes are used.

3. Findings

3.1.Overall Trend

Over the past period, the number of customers increased every month (Figure 1). Although, as seen in Figure 2, the 'pass rate' decreased consistently over the last 5 months.

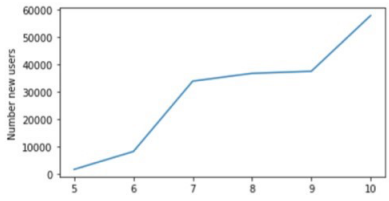

Figure 1. Number new customers

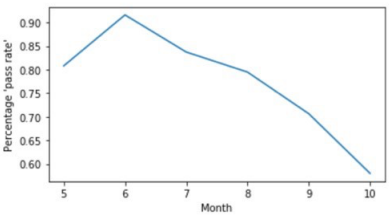

Figure 2. Pass rate

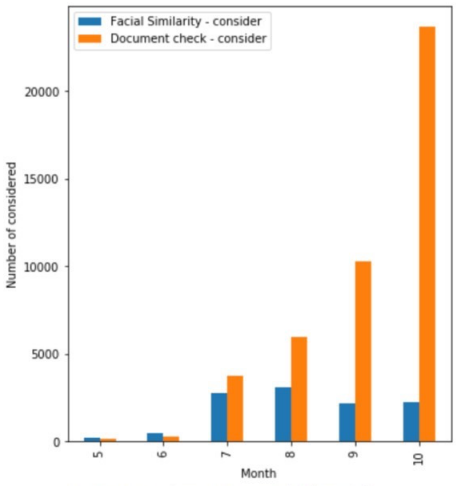

Figure 3. Facial versus Document fail (consider)

So, more and more customers fail the KYC process. When looking into the cause, it can be observed that the problem comes from the Document Check process, with the number of 'consider' almost doubling every month since June (Figure 3), while the Facial Similarity fails decreased over the same period.

3.2.Document Check

It can be observed from the first column in Figure 4 that the result of the document check (result_doc) is highly correlated with the result of the following: image integrity result, image quality result, and conclusive document quality result. The same can be seen in Figure 5 which presents the failure rate of the Document Check by each component.

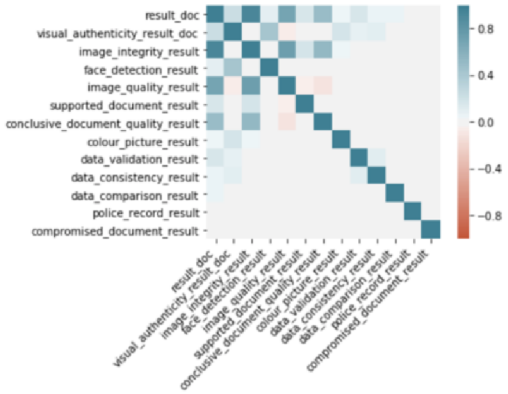

Figure 4. Correlation document check pass/fail with all the document components

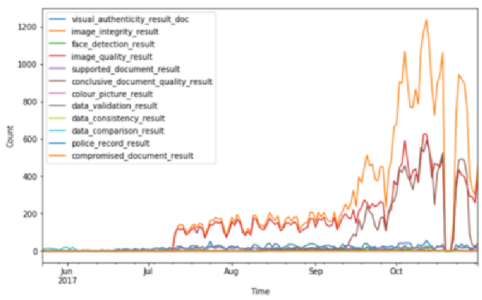

Figure 5. Failure rate for Document Check components

The image integrity asserts whether the fraud inspection passed or not. The conclusive document quality breakdown of the image integrity asserts whether the document was of enough quality to perform a fraud inspection.

Table 1 shows that 90% of document check fails are due to image integrity (fraud inspection fails for a variety of reasons). 30% of document check fails are due to conclusive quality (low-quality pictures – blurred, parts of document not visible)

Table 1. Count of Image Integrity and Conclusive Quality fails

Document Check	Count	Image Integrity Consider	Conclusive Quality Consider
Clear	132,400	0	0
Consider	44,002	39,743 (90%)	13,565 (30%)

3.3.Facial Similarity Check

Although the Facial Similarity Check is not a driver towards the overall failure rate (as seen in figure 3), it is good to understand how its components affect the overall pass/failure rates.

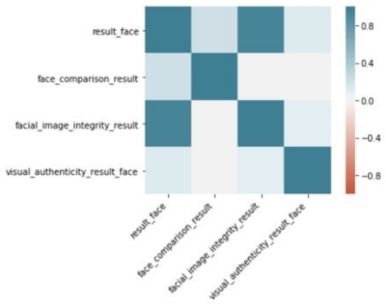

Figure 6. Correlation face similarity pass/fail with all components

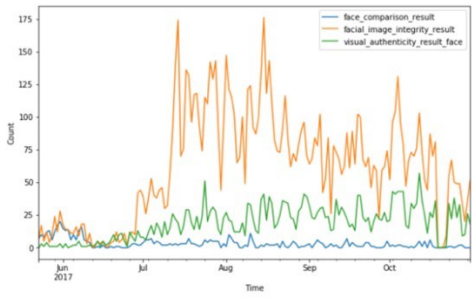

Figure 7. Failure rate for Face Similarity Check components

The biggest driver towards the Facial Similarity Check failure is the facial image integrity result (as observed in Figure 6 and 7). As observed in section 3.1. Overall Trend, the facial similarity check is not a big driver towards the overall failure of the KYC process.

3.4. Document expiration date

It has been observed in some cases that the ID expiration date was before the date on which the customer wished to join X, meaning the IDs were expired. Unexpectedly, sometimes individuals who provided expired IDs were able to pass the KYC process as seen in Table 2. This requires some further investigation into why this issue is present.

Table 2. Number of expired IDs

KYC	Count	Expired IDs	% Expired IDs
Pass	126,889	255	0.2
Fail	49,513	1018	2

3.5.Number of attempts

Each customer has up to 2 attempts to pass the KYC. Although, 1145 customers attempted the process 3 times and 78 attended the process 4 times. On the assumption that the KYC is automatically failed after the second attempt, there are approximately 1300 cases were the process was automatically failed.

4. Conclusions and Recommendation

We have observed a sharp increase in new customers that wish to open a X account. Over the last period the 'pass rate' has consistently decreased, meaning that some problems are present in the KYC process. The main driver towards the failure of the KYC is the document check. 30% of document check failures are attributed to customers uploading lowquality pictures (blurred, parts of the ID are not visible). Furthermore, around 2% of document check failures are attributed to expired IDs.

The following recommendations are proposed:

i) Increasing the quality of the submitted documents

Provide more instructions for customers. On the document submission page clearly explain how to upload good quality ID pictures. Furthermore, include some GOOD versus BAD examples of ID pictures. In the cases when images use sophisticated counterfeiting techniques, or the image is of poor quality, 'expert review' should be considered.

Potential impact: Increase in the overall 'pass rate' from 75% to 83% ii)

ii) Expired documents

Before uploading the picture of the ID into the system, ask the customer to input the expiration date. If the expiration date is lower than the current date, return an error message and don't allow the customer to upload the picture and start the KYC process.

Potential impact: Increase in the overall 'pass rate' from 75% to 77%