NOVA IMS Information Management School

BUSINESS CASES WITH DATA SCIENCE

MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS – MAJOR IN BUSINESS ANALYTICS

Many Gifts UK

Recommendation System

Group T

Ana Amaro, number: m20200598

Filipe Lourenço, number: r20170799

Gonçalo Almeida, number: m20200594

Guilherme Neves, number: r20170749

May, 2021

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

INDEX

1.Business Understanding

1.1. Introduction and Presentation of the Business Objectives

Many Gifts UK is a non-online retailer store based in the United Kingdom that focuses on selling all-occasion gifts not only to end consumers, but also to wholesalers. Recently, the company made the shift to operate fully online and, due to its "loyal" customer base, was able to gather significantly large amounts of data. Now, the company intends to use it to build a recommendation system. In this regard, the main objective Many Gifts UK asked us to work on, was to develop this system, that should not only facilitate their existing customers' purchase path, by providing recommendations of new items at the moment they enter the website, but also be able to offer a solution for new customers. In order to do this effectively, Many Gifts UK also requested a proper deployment plan and a way to measure the quality of the recommendation system provided, which although requested directly from the company's management team, we will not treat as specific business objectives, as they will be "automatically" covered by our use of the CRISP-DM methodology.

Considering the requests presented, we believe that success will be achieved if we are able to create a system that allows Many Gifts UK to increase the amount of new items' sales (by recommending relevant products), while increasing the retention rate of new customers (as they are shown products under their preferences).

1.2. Situation Assessment

To complete this task, we were provided with a single CSV file, containing data regarding the purchases made at Many Gifts UK between 01/12/2010 and 09/12/2011. This accounts for 25,900 valid purchases, across 4,070 different products and 4,372 customers. The information provided was well documented and included: the invoice number, the date, the items (with the product's code, description and price) and quantities acquired, of each product, at each specific purchase, and who was the user making it (identified by means of an ID, while also providing his country of origin).

We agreed with the timeframe purposed to prepare a presentation of 5 minutes to the management team and to deliver a full report regarding the steps followed and the recommendations we want to provide. Moreover, the main costs agreed are the working hours our team will devote to the project and that will be billed to Many Gifts UK in function of the benefits achieved. The main risks associated with this task emerge from Many Gifts UK selling mainly to wholesalers, where a recommendation system may not be as efficient as to an end consumer, but we will try to mitigate this by defining some reasonable assumptions, especially while analyzing the data.

From the terminology point of view, the language used by Many Gifts UK is straight-forward and the only term that may raise some doubts is "wholesaler", that refers to a company that focuses on reselling products to other companies.

1.3. Data Mining Goals

The main data mining goal for this project is to create a recommendation system that is able to have high quality in the suggestions of "new products" made to each user, while being able to provide a relevant solution for the cold start problem. In order to measure the quality of the general recommendation system, we decided the best approach would be to use Mean Average Precision (MAP) at K, being this the measure that we will use to compare across models to select the one delivering the best results. Our goal is to maximize the number of relevant recommendations made to each user (being us only able to suggest K products), while being constrained by the human side of whom we are recommending to, and in this regard we will consider two psychological effects: Serial Positional Effect[1] and Anchoring Effect[2]. The first tells us that people tend to recall better the first and latter items on a list and not as much the middle ones, while the latter states that people tend to pay special importance to the first item/information received (which in this case may refer to more importance/attention being paid to the first few suggestions and/or to the fact that people may make their decision regarding if it is worth or not to use/see the recommendations based on their opinion of the relevance of the first few items). As we are not able to assure that if people lose interest in the middle, they will reach the end, it is of special importance to assure what we can control, this is, the presence of proper results in the first positions. Given this condition, we not only want to penalize false positives (suggest non-interesting products, that waste customers' attention span), but also want to penalize more intensely the false positives that appear in the first positions of our recommendations (where the customer will pay more attention), being this the main advantage of the MAP at K[3].

To measure success, we will compare the performance, in terms of MAP at K (being K = 10, following the industry's standard), of the system developed with a baseline, being that success will be met if the baseline is outperformed by our model. In this case, the baseline will be the simplistic solution of suggesting the most popular items that a specific user has never purchased before. Note that the baseline for a recommendation system, usually, tends to be to suggest the top K most popular items, but as we are interested in building a model that only recommends items never purchased by each user, we decided that conditioning the most popular items to the ones never purchased would provide a fairer and more accurate comparation. Lastly, on what concerns the solution for the cold start problem, we will follow a more subjective criteria of evaluation, assessing by majority vote of our consulting team the theoretical soundness of the solution proposed. The reasoning behind using this more subjective method is related with how we consider that new users are particularly sensible to the recommendations presented, and therefore, our goal is to reach a solution that can be implemented with A/B testing (in comparation with the actual system or lack of it), and assess if it improves the retention rate of new users and the creation of accounts (being this last point a major problem of testing *apriori* a cold start solution objectively, given the possibility of certain users never creating an account and using options like "continue as guest", which don't allow for a proper identification with a fixed ID, leading to less trustworthy results than just measuring it subjectively in a first phase and then using objective metrics with A/B test).

1.4. Project Plan

Concisely, our project will be based on the following steps: 1. Retrieve the dataset, gather some superficial understanding of the data and the variables and try to spot easy to detect problems, by using techniques like statistical analysis of the variables; 2. Deeper exploratory analysis and insight extraction about the data (e.g., analysis of potential problems and reasons behind); 3. Clean the data by acting upon the insight found at step 2 (e.g., solve missing values); 4. Define the model(s) to use, create the conditions to apply it (e.g., split train and test set) and construct it, based on the data analysis done and on the business and data mining goals; 5. Attempt to improve the model(s) performance by tunning the parameters; 6. Verify if the model meets the data mining success criteria, and interpret the results achieved in light of the business context; 7. Present the report to the management team and if the approval is met, deploy the plan.

2.Data Analysis

2.1. Data Understanding

The data collection was straightforward, as all data was contained in a single file, and the information presented in the metadata and business presentation (e.g., variables' meaning and number of records) seemed to be accurate after a first look at the data.

Recurring mainly to visual/graphical assessment, we started by analysing the data, and gathering some insights. The most interesting ones were that: around 65% of Many Gifts' registered customers have made more than 1 purchase; close to 54% of the products were purchased more than 50 times; about 30% of the purchases involved volumes of at least 10 units.

Then, with the use of some visualization tools (e.g., boxplots) and statistical analysis of the dataset, we started to uncover some potential situations of concern in the data. Firstly, there seemed to exist missing values in the field CustomerID and Description. After a closer look at this "situation", it was clear that some of these problems arise from operations that are internal to the company (e.g., "printing smudges/thrown away"), others were actual purchases but where, probably for data collection problems and/or operational reasons (e.g., allow user to purchase without creating an account), there was no customer associated with the transaction, while the remaining involved what looked like "test to the system" transactions, with purchases of products with no description, no customer and price equal to 0. Moreover, we also noticed that some transactions had negative values for the price or quantity. In the price case, there was a simple explanation with the entries with this problem being referent to the operation "Adjust bad debit", while in the negative quantity scenario, the transactions corresponded to returns of products. Still, while exploring the data, we also discovered that some "products" were not actual products, corresponding to descriptions like "Bank Charges" and "Post", being those items easy to identify due to their differently constructed "StockCode" (some of these entries also constituted clear price outliers, for instance, values higher than 10,000 pounds for a "Amazon Fee"). Lastly, we also noticed that some products had more than one price and/or description associated with its StockCode. In this regard, the price change across time seems fairly possible, while for the description case, after a closer look, the majority seem associated with typos, synonymous and other minor adaptations of the same description. Overall, and despite the discussed above, the data quality seems adequate, with a few issues that need to be corrected, and with some data that is not useful for this problem (e.g., non-products), but, nevertheless, containing a well-structured collection of what can be called implicit feedback (in the form of purchases) regarding the users' preferences.

2.2. Data Preparation

Based on the insight reported in the last section, we decided to drop all purchases that were not related with products and that had no customer associated. Considering that non-products cannot be recommended and are mostly related with internal Many Gifts UK's operations, it would be worthless to include those into the recommendation system, and the same can be said about purchases associated with no customers, as it is not possible to construct a "track" of the purchases that lead a user to a specific transaction or that may lead to next products' purchases: even if a customer has made purchases with Many Gifts, if he did not create an account, every time he enters the website, he will be treated as a new customer, as there is no ID in place to identify him.

In what concerns the returns, we decided to not exclude a part of those from our analysis, mainly due to the following reasons: firstly, even if a product is returned, it still reveals interest and shows a interaction between the user and the products (a recommender system that suggested a product, even if returned, would have made a good call in what raises the attention of the user, independently if the product actually delivers the intended value or not); secondly, because we are dealing with a significant number of wholesalers, the dynamic behind the return may be very different from what is considered when dealing with selling to end consumers (e.g., procurement errors or poorly predicted stock fluctuations, that lead to excessive inventory that has to be sent back); and lastly, because, when looking closer at some purchase returns, it is clearly suggested that those seem just "defect items" that were sent back due to that, and not a particular dislike for the product (e.g., large volume purchases of a specific product from time to time with some returned units in the middle). Note that this logic only applies to returns that we do not have a previous purchase of the same product by a specific customer, this is, to returns corresponding to purchases made before the data we were provided with. For the cases we have the purchase and the return registered, in order to not double count that interaction, we decided to drop the return.

At this stage, we also decided to define which products and users we would consider as "needed to be solved" by a cold start solution and which would be used to construct the overall recommendation system. In this regard, any user that has purchased less than 5 different products and any product that has been purchased less than 5 different times was considered as being part of the cold start problem, which allowed for the construction of a less sparse dataset to be fed to the recommendation system. Lastly, in order to apply the recommendation system that we will define in the modelling phase, we will need to further transform the dataset to be able to comply with the required format for each model (which from a purely CRISP-DM methodology should be done in this section as a part of "Format Data"), but we decided to only proceed to this step after the train and test split (mainly due to the need for a "particular" temporal split of data, that is incompatible with being done after the necessary transformation to the data).

3.Modelling

3.1. Modelling Technique and Methodology

As mentioned before, we decided to divide our problem in two different parts: cold start and general recommendation system. Regarding the cold start, we defined that our modelling approach would be based on a combination of a generic recommendation system that suggests the most popular items (POP) and a content-based system. For this particular case, we used the Term Frequency – Inverse Document Frequency (TF-IDF) model[4], and its only requirement is the need to be fed with a single string for each product we intend to have present in the model. Considering we wanted to convey information about the price and the description of the product, we had to first categorize the prices into categories and then combine them into a single string, for each product. Note than when a product code corresponded to diverse descriptions/prices, we used the most frequent to represent it.

For the general model (the one used when a product is purchased at least 5 different times and/or a user has bought at least 5 different items), we will attempt 3 different collaborative-based systems, being those Alternating Least Squares (ALS)[5], Bayesian Personalized Ranking (BPR)[6] and Bilateral Variational Autoencoder (BiVAE)[7].Before explaining the specificities of each model, it is important to clarify what we defined as "preference". To get a measure of preference (in an explicit-feedback problem, it could be seen as the "rating"), and considering we are dealing with a significant part of sales to wholesalers, we decided that using the Quantity in itself would not make sense, as some companies can buy higher amounts than others, due to its size, and not because they actually prefer the product. Being this said, we considered as an interaction-measure/preference, the number of orders where a specific client had purchased a certain product. Note that, as mentioned before, in the cases we have a return but not the corresponding purchase (because it was made prior to that data we have available), we will count the return as an interaction, capturing indirectly the purchase to which that return corresponds to. Regarding the models used, both the BPR and the BiVAE required us to format the dataset in order to have each line reflecting an interaction between a client and a product, having the CustomerID, StockCode and the preference index (that we defined as the number of orders where the user buys the specific product). When it comes to the ALS, the model requires data to be in an item-user matrix, meaning that each row represents a product and each column a user, while the entries correspond also to the preference.

In order to properly evaluate the collaborative-based system we want to implement, we needed to create a proper test design: we split the data at a temporal level, reserving around 74.1% of the data (which corresponds to the first 46 weeks of data) for training our model and the remaining for testing (the last 7 weeks). In a nutshell, what this split will do is to mask purchases of customers after the week number 46, so that our model only finds patterns until that week and then attempts to predict the remaining 7 weeks' purchases using the knowledge gathered during the training phase.

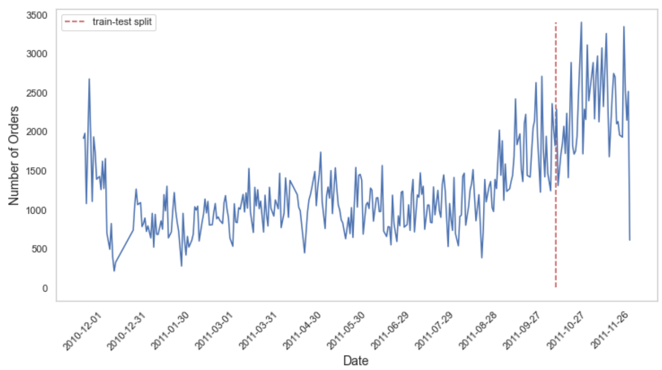

Figure 1 – Division (red line) between train and test set, made with a temporal base.

3.2. Model Construction and Assessment

With the needed adjustments to the data for the diverse models applied, we moved to their implementation, while combining different parameters to optimize the results achieved. In this regard, we present in a table below the values and parameters we attempted to tune (we used bold to highlight the parameters that achieved the best score after the optimization process), being that those were selected after considering (and in some cases following) the implementation's default values, the suggested values by the authors of the models (when available), previously used values across the industry to optimize recommendation systems and our own experience with the selected models. Note that, due to limitation of space and better readability, we decided to leave the meaning of each parameter out of the present report, but in the Jupyter Notebook accompanying this report (before each optimization), we include a brief description of what each parameter means. Finally, it is also relevant to point out that although, and as explained in section 1.3., the main factor used to compare models' performance was the Mean Average Precision at K, we still kept under close attention the metrics Precision at K, Normalized Discounted Cumulative Gains (which in some ways attempts to capture the same as the MAP at K) and the Recall at K.

Model	Parameters
ALS	Alpha: 1.5, **3**, 15, 40; Regularization: **0.05**, 0.01, 0.001; Latent Dimensions (Factors): **50**, 150, 200; Iterations: 200, **500**
BPR	Latent Dimensions (K): 50, 75, 100, **125**; Maximum Interactions (Max_iter): **250**, 500; Learning Rate: 0.001, **0.01**; Lambda Regularization: **0.001**, 0.0001
BiVAE	Latent Dimensions (K): 10, **50**, 100; Encoder Structure: 20, **50**; Activation Function (Act_fn): "**tanh**"; Likelihood: "**pois**"; Number of Epochs (N_epochs): **100**; Batch Size: **128**; Learning Rate: 0.01, **0.001**; Beta (beta_kl): 1, **1.5**

Table 1 – Parameters' values used for optimizing the different models, with the best selection for each in bold.

Still in the construction phase of our models, we implemented a simple model that returns the most popular items, with two variations: one only returning products never purchased by a particular user (to use as benchmark against the best optimized model constructed) and a second version that returns the most popular items without any consideration by the user being passed to the model or even if there is no user being passed (this will be used as part of our cold start solution). Moreover, we also proceeded to the construction of the content-based model, that due to its nature did not require specific tunning of parameters, except for the definition of a criteria of similarity between the strings passed. In this regard, we selected the cosine similarity.

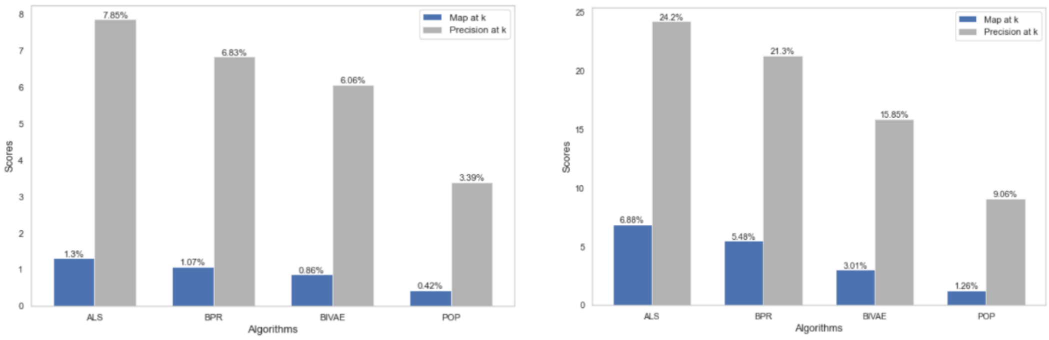

Figure 2 – Map and Precision at K, in the left recommendations for only products never purchased, and in the right for all products.

With the parameter tunning phase done, the collaborative-based model that gave the best score was the ALS model. This model was able to achieve a MAP at 10 (K) of 1.3%, as visible above, surpassing by far the benchmark score given by recommending the most popular products never purchased by a specific user. Furthermore, in terms of Precision at 10 (K), we also achieved an interesting score of 7.85%, which given the easy interpretability, we also decide to present here. Moreover, and although not asked, considering we are dealing with a significant amount of wholesalers, and due to the repetitive purchase behaviour demonstrated by them, we decided to also (with no specific optimization, using the best parameters achieved for the recommendation of new products, only as a proof of concept) present the results achieved for a scenario where for convenience Many Gifts UK decides not only to show new products, but also attempt to facilitate the costumers' purchase path by suggesting any item, including already purchased ones. For that particular case, as expected, the results are significantly better, being worth to point out that suggesting the most popular products (POP), independently of previous purchases, was, in terms of MAP at 10 (K), not only surpassed by our models created with this same condition (possibility to suggest already purchased items), but was also exceeded, by a slim margin, by our ALS model only suggesting new items, what given the context of repeated purchases (wholesaler environment) is a very remarkable result.

Overall, we believe to have been able to achieve success in what concerns the data mining goals defined. Firstly, we were able to construct a model that performed better, in MAP at 10 (K), than the benchmark defined by recommending the most popular items never seen by each user (with a score more than 3 times better). Furthermore, we took advantage of the properties of content-based models to find a solution that our team considered theoretically valid for the cold-start problem, that we will later elaborate on, and how it can be properly implemented in conjugation with the collaborative-based model.

4.Evaluation

4.1. Evaluation Results

Being the model construction done, and having we achieved data mining success with this project, it is time to evaluate how was the performance in terms of business goals.

The model created is able to offer relevant suggestions 7.85% of the times from the 10 suggestions (at K), while prioritizing the order of the suggestion being passed to the customer (using the MAP at 10 (K) as the optimization metric), which although not sounding particularly impressive, achieves a significant improvement over the simplistic solution of suggesting the most popular items never purchased by a certain user. Despite the good comparative performance, we believe that one of the particular limitations to the score achieved was the fact that dealing with wholesalers is significantly different than dealing with end customers, where the wholesalers tend to make more frequently repeated purchases (probably of the products they sell the most), potentially showing a low risk tolerance: when an end customer purchases a new product, the worst scenario for the customer is to have lost the money paid for it, while a wholesaler, when purchasing a new product, has to consider things like the impact on its brand (e.g., Is this product complying with our company's values or image?) and the risk of ordering too much (if the customers do not like, they end up with excess inventory) or too little (if in fact that product is appropriate, customers may be upset by the lack of inventory and may go to competitors). With this in mind, and although in the business presentation (and as stated at 1.1.) was clear that Many Gifts UK wanted a system to suggest new products to customers, given the low resources involved (building on top of what was constructed), we decided to also offer a tool that instead of focusing on new items is focused on convenience, providing to customers recommendations even of items already purchased by them, which as mentioned, allows for a significantly higher recommendation's success rate. Either way, we believe to have developed a model that when put in place will allow to achieve the first business success criteria of increasing the sale of new items, by providing the most relevant suggestions to Many Gifts UK' customers.

Regarding the second business goal, we will explain on the Deployment section how we intend to extract the most out of the combination of the recommendation of popular items with the content-based recommendation system, but overall, we believe to have been able to face properly the challenges posed by new users, and how susceptible they are to proper recommendations. Being this said, we believe that our recommendations will help guide new users in their purchasing journey up to the point we have captured their behaviour and include them on the collaborative-based system, which seems the proper strategy to increase the retention rate, in the future, as defined in the business success criteria. Building on the solution found to this business need, particularly on the use of the TF-IDF system, we found some significant benefits to the business, which expand far from only providing valid recommendation to new users. Firstly, TF-IDF, by recommending not based on products' popularity, but based on similarity between products, allows to properly recommend new lines of products that Many Gifts UK may introduce, as well as less popular products (that usually get trapped in a vicious cycle, as they are never recommended, they never get purchased, which makes them not being recommended and so on). This will allow for these products to be recommended up to the point they reach the threshold of 5 purchases so they can become part of the ALS model (after re-train). Moreover, we also believe that this content-based system can be useful to recommend similar products in case a user enters a page of a product that is out of stock, allowing the user to probably find a substitute, and not risking losing it to the competition.

Before proceeding to the Review of the Process, we would like to make some suggestions regarding future endeavours. Firstly, we believe that the process of creating a description for products should be automatized somehow in order to avoid the systematic different descriptions found to the same product. Moreover, it would be good to be provided with the recommended selling price, instead of the price applied in the moment of the purchase, as we have realized that there are variations that may have to do with price discounts and this sort of activities. Lastly, and as a major suggestion for better applicability of recommendation systems, in the future, we believe that Many Gifts UK could and should implement a rating system, even a simple one with a like and dislike button in each product. This would allow for a later endeavour to use explicit feedback, which provides a significantly higher degree of certainty regarding the true preferences of customers, avoiding therefore the assumptions that we had to make over the current project.

4.2. Review Process

After a careful review of the process and steps undertaken this far, our team did not spot significant mistakes or failures. Despite this, and as reported, we had to deviate partially from the CRISP-DM methodology, as we were not able to fully prepare our data before the Modelling phase (the methodology is unclear on when this model specific data preparation should be done, but a brief mention to it is made in the Format Data section). Being this said, we believe that we used the methodology as rigorously as possible, with the needed adjustments that are case specific. Moreover, the process followed complied with the initially defined in section 1.4. of the current report and our team made sure to evaluate the process followed not only in light of the data mining and business goals' context, but also from a theoretically quality assurance.

4.3. Determining Next Steps

At this stage, and although we could implement different modelling techniques, try to optimize with more parameters the models used, and/or even reconsider the value we initially defined to make our interaction matrix less sparse (by attempting different values), we are confident that the solution proposed presents a proper answer to Many Gifts UK's problem. Furthermore, any other move than deployment, would demand a larger time frame (and potentially the expansion of computational power, as the optimization phase was already fairly time costly), which seems an unnecessary expense, given the opportunity of validating the assumptions made with the release of an already viable solution. We suggest therefore the deployment of this project.

5.Deployment

5.1. Deployment Plan

Regarding the strategy for deploying, and as previously mentioned, we believe that a system with a simple like/dislike approach would not only allow for more precise results, but also for quicker feedback of the quality of the recommendations made. In this regard, in an optimal environment, the strategies laid here should be seen as a temporary endeavour while this improved system is put in place, as it will take time for the likes/dislikes to achieve critical mass. Note that even with the explicit feedback in place, the solution provided for the cold start problem may still be the best option for this system with some potential adaptations (e.g., instead of combining the content-based model with the most popular items, we could replace the latter with the most "liked" products).

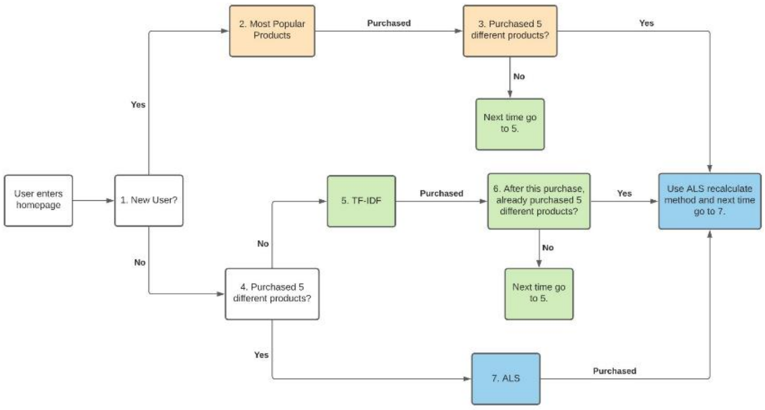

Figure 3 – Schema of the suggested Deployment Plan.

Being this said, to implement our project, the first step would be to create an automatized pipeline to follow the cleaning and preparation steps done in the Jupyter Notebook delivered together with this report (particularly removing non interesting transactions). Then, and according to the schema above (and the prototype script here written in Python), we should put in place an automatized way to move properly the customers in their "journey" with Many Gifts UK, in such a way that after each purchase made, the recommendation system is updated accordingly: Once a user logins at the website, we would evaluate if he already made a purchase or not, if he never made a purchase, we would recommend the most popular products (as we have no information about it, the best guess is that he will be an average customer). In case a purchase was already made, but with less than 5 different products purchased, we would suggest him new products based on the related "likeness" with previously purchased products, using TF-IDF. Lastly, if the user already achieved 5 different products purchased, we would use the ALS system. At the end of each purchasing experience, we would either pass the user for the "recalculate" method of the ALS implementation used (in order to either update user preferences or insert him for the first time in this model) or find the appropriate solution based on the number of different products purchased. Considering the possible computational costs associated with this endeavour, and not knowing if Many Gifts UK can support those, there may exist the need to delay the recalculation of the ALS, not after each purchase, but after a given number of purchases, with the risk of suggesting to customers repeated and/or not as relevant products for a short period of time. Lastly, we decided to launch a static model, available in this dashboard, where Many Gifts UK can see the recommendations we would make to their users, at this point in time, according to the solutions proposed.

5.2. Plan Monitoring and Maintenance

Given the importance of this endeavour in the day-to-day operations of Many Gifts UK, it will be of extreme relevance to keep track of the system's results and keep it running in the proper environment. Being this said, and although the method "recalculate" provided in ALS's implementation helps to quickly update preferences or include customers once they buy 5 different items, we should create an automatized system that after a specific number of purchases re-evaluates the MAP at K (and potentially the Precision at K), in order to determine if the model is not producing satisfactory results anymore, according to a threshold. If this is the case, this should activate a trigger in the system that should re-train the full model or even try to find out if any attempted model has become better than the ALS (or at least this ALS's parameterization). Note that a re-train solution based on a temporal schedule is highly not advisable, as that does not capture the potential growth of the business and may create a too significant delay between trainings. Lastly, Many Gifts UK should keep close attention to the retention rate of new users (until reaching the 5 different products purchased) using A/B testing between the (lack of) current solution and the one proposed by us, to be able to compare effectively if the solution proposed for the cold start problem is producing the desired outcomes.

5.3. Conclusions and Brief Review of the Project

Although the review of the project should be discussed with the various stakeholders - in order not only to identify possible pitfalls or processes that might be improved, but also to keep track of the results of the deployment plan and if some major deviance occurred -, we believe that we can already present some insight in the form of general conclusions of the project. Overall, we believe to have been able to provide Many Gifts UK with a model that is able to make proper recommendations for their customer base, while also being able to accommodate new users, aiming therefore at solving the "cold start" problem. Moreover, we believe to have constructed a detailed and well-planned deployment plan, that if followed will contribute to the perpetuation of the benefits reached, while creating the structure for future endeavours to build on.

6.References

[1] Troyer A.K. (2011). *Serial Position Effect*. Encyclopedia of Clinical Neuropsychology. Springer, New York, NY. Retrieved from: https://doi.org/10.1007/978-0-387-79948-3_2232

[2] Program on Negotiation – Harvard Law School. *What is the Anchoring Effect?* Retrieved from: https://www.pon.harvard.edu/tag/anchoring-effect/

[3] Tan, R.J. (2019, March 24). *Breaking Down Mean Average Precision (mAP)*. Retrieved from: https://towardsdatascience.com/breaking-down-mean-average-precision-map-ae462f623a52

[4] Jalilifard, A., Caridá V.F., Mansano, R.S.C. & Fonseca, F.P.C. (2021, January 25). *Semantic Sensitive TF-IDF to Determine Word Relevance in Documents.*

[5] Hu, Y., Koren, Y., & Volinsky, C. (2008). *Collaborative filtering for implicit feedback datasets*.

[6] Rendle, S., Freudenthaler, C., Gantner, Z. & Schmidt-Thieme, L. (2009). *BPR: Bayesian Personalized Ranking from Implicit Feedback*.

[7] Truong, Q.T., Salah, A. & Lauw, H.W. (2021, March). *Bilateral Variational Autoencoder for Collaborative Filtering*.