

Subscription Models and Analytics in E-Commerce

Rohan Mayya¹, Pavan Puranik², Romil Balar³, Saifur Rahman⁴ and Dr. Selva Kumar S⁵

¹⁻⁵BMS College of Engineering, Bengaluru, India

Email: rohanmayya.cs17@bmsce.ac.in, {pavang.cs17,romilbalar.cs17,saifur.cs17,selva.cse}@bmsce.ac.in

Abstract—The subscription model is a rapidly emerging business model in which products are delivered to customers at regular intervals instead of multiple independent purchases. Based on all of this, our goal is to build a feature-rich Shopify app that includes features like delivery date handling, product add-on for additional sale, and onion analysis. and basic customer retargeting mechanism. Machine learning for special businesses registers based on customer behaviour, we want to predict whether the customer is aroused or not. We are looking to use metrics like number of purchases, delivery time per purchase, months of subscription opted for, whether other products were brought, other regular orders placed before opting for subscription.

Index Terms—shopify, shopify app, subscription model, subscription-box model.

I. INTRODUCTION

The rise of global internet began in the late 1980s to early 1990s as it started moving into the limelight for public usage rather than being restricted to government bodies and agencies. The evolution of modern day e-commerce through the internet as an idea started with the advent of the world wide web as a concept and a few years later, the emergence of social media. It started with Viaweb, founded by Paul Graham, who later sold to Yahoo! for \$49 million. This idea soon started gaining traction among small business owners and entrepreneurs looking at starting up. Here was a platform which could enable a level of communication wherein people could directly communicate what they wanted to buy/sell to anyone in the world. This was cue for business across the world to reach out to a much larger consumer base rather than just the few people who would physically come to wherever their store was located. After that, Shopify emerged when a few entrepreneurs wanted an online store for snowboarding equipment. Dissatisfied with the existing e-commerce products on the market, they built their own. They soon launched the platform and named it Shopify. After a rocky start, Shopify has now captured \$123bn of value and stands as the No.1 Canadian company (in terms of valuation). The motive is clear: everyone who aspires to start an E-commerce company is looking to outsource the heavy lifting. Payment gateways, product listing, inventory stock: these are all problems that are much rather relegated to a third party system. This also solves the problem of people not having the technical expertise to build and setup an online store by themselves.

There is another segment of E-commerce that has been booming of late: Subscription models. Subscription models are a fast emerging business model wherein products are delivered (usually through packages/boxes) to the customer at regular periods of time instead of multiple standalone purchases. The idea is to not have to order the same thing repeatedly and also involve a certain amount of customization that lets the vendor send special items usually as part of the package. Subscription works incredibly well for all kinds of daily use products as

well as certain niche requirements that the user has. In the United States alone, back in 2018, subscriptions were estimated to be around \$15bn in value. This is a market that has seen blazing-fast growth year on year. Many of the listed e-commerce platforms have an out-of-the-box solution to handle subscriptions, as it consists mostly of recurring payments, but the literature survey we've carried out deems that they are not flexible enough. Churn rate is a measure of the number of customers who move out of the sales funnel at a particular stage. Simply put, a low churn rate is an indication that the business is doing really well in retaining customers who have purchased the product in the past.

II. SUBSCRIPTION MODELS

Subscription models are a fast emerging business model wherein products are delivered (usually through packages/boxes) to the customer at regular periods of time instead of multiple standalone purchases. The idea is to not have to order the same thing repeatedly and also involve a certain amount of customization that lets the vendor send special items usually as part of the package. Subscription works incredibly well for all kinds of daily use products as well as certain niche requirements that the user has.

In the United States alone, back in 2018, subscriptions were estimated to be around \$15bn in value. This is a market that has seen blazing-fast growth year on year. Many of the listed e-commerce platforms have an out-of-the-box solution to handle subscriptions, as it consists mostly of recurring payments.

III. REVIEW OF PAST WORK

Existing systems have a tremendous number of flaws. From our extensive research: The biggest one being the customization of the subscription order. There is no leeway to defer your order's date, no way to upsell products that you would otherwise like to upsell, and no control over pausing and resuming orders work. Existing machine learning models that have been deployed for customer retention do not fare well when it comes to important metrics that define purchases made on a subscription basis. Bold Subscriptions isn't API-first, which makes their platform extremely lacklustre in our eyes. Some of their features are gate-kept behind a \$200 paywall, that a developer could instantly implement if they had access to their APIs. ReCharge on the other hand is a lot better, is API-first, and thinks of the end-user, and provides flexibility. But they fall short in the feature department. As mentioned earlier, things like a lack of control over pausing and resuming a subscription for a customer and a lack of selection of delivery dates/product options really hurt the experience.

IV. MACHINE LEARNING APPROACH TO CUSTOMER RETENTION

Based on customer behaviour, we would like to predict if the customer is going to be churned or not. This boils down to a binary classification problem, and will give us results between 0 and 1, 1 being most likely to be churned. Once we triangulate those users who are likely to be churned, we can draw similarities between this user pool. These similarities/ user parameters can then be used for our customer retention analysis. If such parameters are detected in future users, a different approach may be employed towards them in an effort to reduce churn. The success of these methods being employed can be used to further validate our prediction model.

For each model we perform the following:

Parameter Identification: We discover the parameters we need in each model we use.

Finding Best Model: After finding required parameters, we train the model using the training data set.

Performance check: After model training, we will evaluate the model's performance using the test set.

1) Logistic regression classifier: It creates a model that explains how independent variables contribute to a binary dependent variable. In other words, it defines an equation which works on features which are predicted to impact churn, using which it tries to estimate the coefficient for each user (based on churn).

The logistic regression classifier used in the telecom domain, produced an ROC of 0.83 [1].

2) Support Vector Machines: this algorithm creates an n-dimensional space where 'n' is the number of features used, with each user represented as a point in that space. The goal is to separate these points into 2 distinct groups, with a good amount of separation between these groups. This is done by finding a hyperplane. The SVM model makes a prediction for each data point for whether the user will churn or not. Standard SVM is not suitable to tackle this problem. Thus, we introduce the "Small Sphere and Large Margin machine" specifically for customer churn prediction. It is shown to achieve the best performance in experiments and is proven by experimental results [2].

3) Random Forest Classifier: Random Forest is an ensemble-based algorithm. It is composed of a large number of randomized decision trees. Each of these trees will classify a point, and by the majority vote, the final decision is reached. This technique helps tackle over-fitting. The ROC of standard RF and RF with data pre-processing has proven to be much higher than the aforementioned techniques, on a telecom data-set, for customer churn. [3]. Use of ROC curves as a performance metric: We can investigate the true positive and false positive rates using the ROC Curve. We can also calculate the area under the curve and use it as a metric of the prediction power, hence performance of our model. A value closer to 1 says that the model differentiating the test sample into two classes very well.

Feature Importance Based on our analysis we will determine which features impact churn. For this we will use the concept of variable importance, that is to measure how impactful each variable is for our model. This analysis forms the crux of our system. We can identify the features that make a customer more likely to churn and plan targeted strategies to avoid said churn. This feature finding gives our customer retention analysis direct and tangible results which make sense to the system-user.

V. PROPOSED SYSTEM

Build a Shopify app that leverages the latest tech to build a compromise between ReCharge and Bold Subscriptions. Offer high-level APIs and more control over what features can and cannot be built. Take a “headless” approach as far as possible where full control resides in the customers’ hands, not the other way around. Deploy machine learning models that captures all the relevant features and gives the business owner insight as to whether a particular customer is likely to keep purchasing and if no, tailor their value delivery/offerings accordingly

VI. HIGH LEVEL DESIGN

Shopify is very powerful in terms of the services that it offers and that can be related to its architecture and the compliance policies that gives a developer very few ways to interact with their systems while still offering a lot of control through the concepts of Shopify Apps.

So, Shopify also makes it convenient for the developers to add services/features on top of what Shopify is capable of. We will be taking advantage of this feature to build a Subscription Box Model on top of a generic eCommerce platform. The central server/backend server has many interfaces towards multiple services that it interacts with. In simple terms, if we talk about order placement for a subscription box product, it follows the following chain of events:

The product which is to be treated as a Subscription is already marked as Subscription through “product type attribute”, which makes it convenient for the server to distinguish a Subscription product from a regular product. A single order on Shopify can contain multiple items, which makes it difficult for us to accurately find the items which are marked as “Subscription”. However, we can iterate over the list of items and quantities to find out all the information required to create a database entry on our system for the subscription items in the order. When an order is placed on Shopify, the same order data is sent to our server through a webhook.

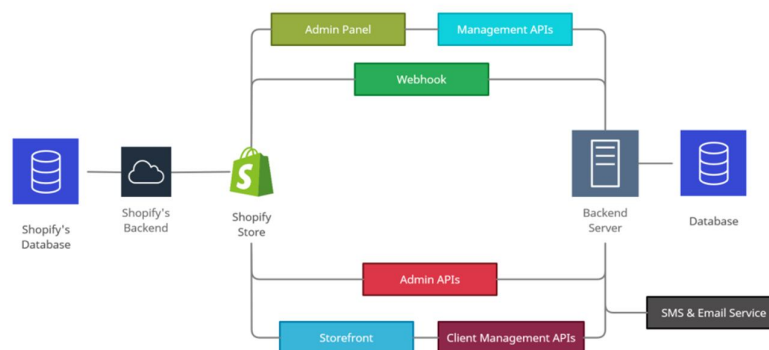


Figure 1. High Level Design

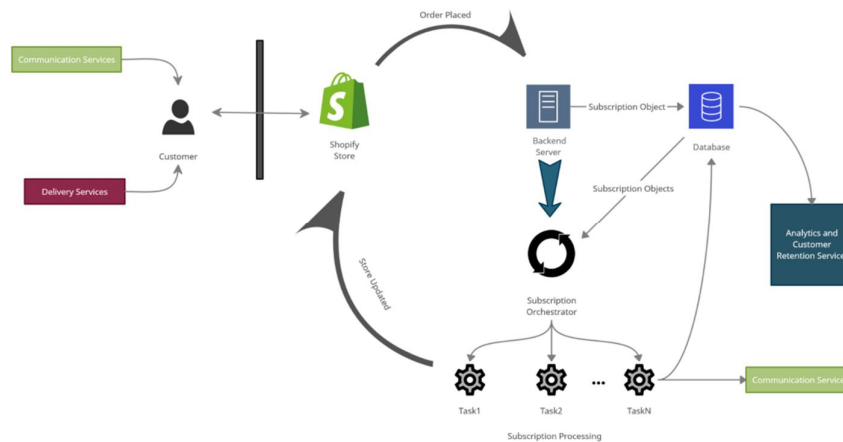


Figure 2. System Architecture

Webhooks are a powerful tool for apps that want to stay in sync with Shopify or execute code after a specific event occurs in a shop. The webhook payload is used to find out all the required items to be processed as a Subscription product/item. Then the information is saved on to the database using a flexible MongoDB schema. The Subscription Orchestrator's job is to process the subscription items on a daily basis. This task is time consuming and requires SLAs to be met and thus would require us to split the amount of data we have to process into multiple batches depending upon the number of subscriptions to be processed (customer base).

Multiple jobs/tasks can be created and delegated to temporary subsystems. The system should be scalable enough to finish the task in a given amount of time irrespective of the number of the data points to be processed. The task executor thus needs to interact with many other services including Shopify's API, communication services and the database. The task executor's job is to process the subscription objects according to its current state, which includes placement of an order on shopify, skipping and calculating dates, scheduling SMSes and Emails to the customers, updating the subscription objects and saving it back in the database. The orders placed by the executor can then be used by the Store Owner/Inventory Manager/Staff to deliver the items to the customers.

The task executor's job is to process the subscription objects according to its current state, which includes placement of an order on shopify, skipping and calculating dates, scheduling SMSes and Emails to the customers, updating the subscription objects and saving it back in the database.

The orders placed by the executor can then be used by the Store Owner/Inventory Manager/Staff to deliver the items to the customers.

VII. CONCLUSION

We would like to actively disrupt the subscriptions space by making the end experience better for all customers. By offering a tremendous amount of low level control, an extensible API, and a set of core features, we believe we can deliver a better experience than the competition. We also believe that we can deliver very accurate results in terms of whether a customer is likely to churn or not based on the unique features that we will identify based on customer activity on the site. Over time, we would also like to dive deep into the subscription space, and actively apply principles learnt in making subscription boxes to regular purchases, as the retention rate is generally much higher.

One of the things we would love to explore is the possibility of setting up recurring payments on the platform so that any person that opts for a subscription based purchase can very simply enter their PIN and authorize payments in one click rather than manually having to make payments at regular intervals.

ACKNOWLEDGMENT

Our team would like to express our deepest and sincere appreciation to all those who have provided us the opportunity to complete phase 1 of our project. We would like to offer our special gratitude to our final year project guide, Dr. Selva Kumar, Assistant Professor, Department of Computer Science and Engineering, B.M.S.

College of Engineering, Bengaluru, whose contribution in helping us venture in the tedious yet intriguing tasks of the project, by providing us with suggestions and encouragement, has helped us to find the starting point for our project especially in coming up with design and documentation.

We would also like to acknowledge with much appreciation the crucial role of the head of the department, Dr. Umadevi V, Department of Computer Science and Engineering, B.M.S. College of Engineering, who has invested her full effort in providing us the support required in achieving the goal. We have to appreciate the guidance given by other supervisors as well as the panels especially in our project presentations that have improved our presentation skills thanks to their comments and advice.

We would like to express gratitude towards our guide and Department of Computer Science and Engineering for their kind cooperation and encouragement which helped us move ahead in the project implementation.

REFERENCES

- [1] B. N. K. Sai and T. Sasikala, "Predictive Analysis and Modeling of Customer Churn in Telecom using Machine Learning Technique," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 6-11, doi: 10.1109/ICOEI.2019.8862625.
- [2] Y. Xu, "Predicting customer churn with extended one-class support vector machine," 2012 8th International Conference on Natural Computation, Chongqing, 2012, pp. 97-100, doi: 10.1109/ICNC.2012.6234646.
- [3] P. Swetha, S. Usha and S. Vijayanand, "Evaluation of Churn Rate Using Modified Random Forest Technique in Telecom Industry," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), Bangalore, India, 2018, pp. 2492-2497, doi: 10.1109/RTEICT42901.2018.9012251.
- [4] Ramakrishna Vadakattu ; Bibek Panda ; Swarnim Narayan ; Harshal Godhia "Enterprise subscription churn prediction", IEEE International Conference on Big Data (Big Data), 2015
- [5] M.C. Mozer ; R. Wolniewicz ; D.B. Grimes ; E. Johnson ; H. Kaushansky. "Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunication Industry". IEEE Transactions on Neural Networks, Volume: 11, Issue: 3, 2000.
- [6] Andrew h. Karp, using logistic regression to predict customer retention, 1998.
- [7] Guoxun Wang , Liang Liu , Yi Peng , Guangli Nie, Gang Kou, G., and Yong Shi. "Predicting credit card holder churn in banks of China using data mining and MCDM". IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2010.
- [8] Adnan Idris, Asifullah Khan, yeon soo Lee, "Genetic programming and adaboosting based churn prediction for telecom. In the proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1328-1332, 2012.