

PROFILAGE DES CLIENTS PAR LES TECHNIQUES DE PROCESS MINING

Hippolite KABANGO

DOI: <https://doi.org/10.5281/zenodo.13923005>

Published Date: 12-October-2024

Résumé:

Dans un environnement aussi concurrentiel comme le commerce, le suivi des activités de sa clientèle devient une activité continue pour les organisations et entreprises à vocation commerciale. Avec le développement du web, la virtualisation des marchés est devenue un défi pour les différents marchands.

La complexité de l'activité due au nombre des clients et à leur cartographie rend la tâche très difficile dans la compréhension des besoins de sa clientèle et de s'y adapter enfin de capitaliser le bénéfice. Au-delà des techniques ordinaires d'analyse des données, le process mining constitue une approche par excellence dans l'extraction de l'intelligence cachée par les traces des utilisateurs enregistrées dans le fichier journal du système d'information.

Par les habitudes on peut se servir des techniques de process mining pour comprendre la routine des utilisateurs ou clients, identifier le goulot d'étranglement et améliorer l'expérience client.

Abstract:

In such a competitive environment as commerce, monitoring the activities of its customers becomes a continuous activity for organizations and companies with a commercial vocation. With the development of the web, the virtualization of markets has become a challenge for the various merchants.

The complexity of the activity due to the number of customers and their mapping makes the task very difficult in understanding the needs of its customers and finally adapting to capitalize on the profit. Beyond ordinary data analysis techniques, process mining is an excellent approach in extracting the intelligence hidden by the traces of users recorded in the log file of the information system.

Through habits, process mining techniques can be used to understand the routine of users or customers, identify the bottleneck and improve the customer experience.

Mots clé: Profilage, Process Mining, Client, Hippolite, Kabango.

1. INTRODUCTION

Le profilage est une technique marketing qui permet aux entreprises de connaître leurs clients et pouvoir adapter l'offre (service et produits) à ces derniers. Cet exercice se base sur l'identification systématique des caractéristiques et aspirations des clients en vue d'un service personnalisé. Il sied de relever ici qu'au moyen de ces profils, les entreprises peuvent optimiser leurs prestations en leur proposant soit des services soit des produits, répondant directement aux intérêts des clients, et additionnellement faciliter une communication efficace car ayant la maîtrise de leurs habitudes.

Partant de cet aperçu général, le profilage sur le plan marketing consiste à regrouper l'ensemble des caractéristiques démographiques, géographiques et psychographiques comme motivations habitudes d'achat. Il permet à l'entreprise de se focaliser sur sa cible principale, d'optimiser ses efforts commerciaux et d'augmenter ses ventes.

L'optimisation de cet exercice est tributaire à l'amélioration des outils utilisés dans le profilage. Avec les techniques du process mining, nous pouvons arriver à profiler un client pas simplement sur base de ce qu'il est, où il habite mais également sur base de ce qu'il fait (comportement) sur base des données factuelles enregistrées dans le fichier journal du système d'information (Event Logs) et non des hypothèses.

Etant donné que les activités dans le domaine digital sont gérées par un système d'information, tout ce qu'on réalise ou on tente de réaliser, laisse toujours des traces qui sont identifiables, compréhensibles et interprétables par l'exploitation des outils appropriés.

Le profilage dans notre contexte, consiste donc à recenser les profils des utilisateurs d'une plateforme e-commerce sur base de leur comportement c'est-à-dire, étant donné que toute procédure d'achat en ligne est tributaire de l'exécution d'un certain nombre d'activités, profiler veut tout simplement dire, grouper les clients en terme de capacité d'exécution des activités du processus et dire pourquoi tel s'est arrêté à tel ou tel autre niveau tout en déterminant leur degré de complétude.

1. Technique de profilage

Partant de sa position intermédiaire du Process Mining ou Fouille de processus entre l'informatique décisionnelle (Business Intelligence) et la fouille de données (Data Mining) d'une part, et la modélisation et l'analyse de processus métier d'autre part, ce dernier permet de découvrir, superviser et améliorer les processus réels (non définis à priori ou non supposés) par l'extraction des connaissances à partir des traces d'exécution (event logs) facilement disponibles dans le système d'informations courant, il reste la technique la mieux indiquée dans la constitution du profil sur base des données factuelles.

2. Définition des concepts

✓ **Activités (événement)** : est un ensemble des tâches exécutées dans un contexte durant un intervalle de temps bien déterminé pour un but précis. Elle est caractérisée par des attributs qui contiennent des données associées à l'activité¹.

✓ **Journal d'événements (logs)** : Un journal d'événements est un ensemble des instances de processus. Il contient plusieurs informations sur les processus donnés. Les journaux d'événements sont le point de départ pour l'exploration des processus.

3. Motivation

Nous avons choisi le process mining dans le profilage des clients dans un environnement e-commerce au regard de son objectivité, sa complétude et sa rapidité.

- Objectivité : il est basé sur les données réelles et non sur les hypothèses. L'exploration (discovery) de ces données dans le fichier journal du système d'information et elles sont directement liées aux activités qui se sont réellement passées par le Alpha algorithme ;
- Complétude : il utilise toutes les données des processus enregistrées, y compris les exceptions et les variantes inconnues ;
- Rapidité : par l'automatisation des actions qui concourent à la production des résultats.

4. Techniques du Process Mining

Les techniques du process mining sont perçues et présentées en ces trois catégories d'algorithmes :

- La découverte : ces algorithmes consistent à explorer les journaux des événements dans le but de découvrir les traces laissés par les utilisateurs dans leur utilisation de la plateforme, puis définir un processus autour de ces activités ;
- La vérification de la conformité : à partir du moment où nos processus sont définis, l'étape qui suit est de s'assurer que les utilisateurs l'applique correctement et de vérifier que les processus fonctionnent comme prévu. La vérification de conformité consiste à analyser les journaux des événements pour s'assurer que les actions sont alignées sur le modèle de processus tel que prévu.
- L'amélioration : ici, il s'agit de déceler les écarts entre le modèle de processus prévu et celui exécuté en vue de l'amélioration continue.

Ainsi, dans le présent travail, nous allons nous focaliser à la découverte de processus par l'utilisation d'un algorithme de découverte appelé Alpha algorithm.

¹ B. ABDERRAHIM, E.G. ANASS, Etude des techniques de fouille de données « Process Mining », ENSIAS 2019, p18

5. Outils du Process Mining

Les outils dont il est question ici, c'est sont les plateformes mise en place pour créer les modèles qui seront compréhensible dans l'interprétation et découverte des activités manquantes ou du goulot d'étranglement. En ce qui nous concerne nous avons :

- PROM : fournit un environnement polyvalent et extensible pour l'exploration de processus. Il fournit des plug-ins permettant d'extraire différents types de modèles à partir des journaux d'événements, par exemple, la construction d'un processus et de modèles organisationnels. De plus, il prend en charge la conversion et l'analyse de modèles. En utilisant des techniques de contrôle de conformité, les modèles peuvent également être comparés à la réalité et les modèles existants peuvent être améliorés avec des informations supplémentaires ;
- Disco : un outil complet de fouille de processus développé par fluxicon qui nous permet de faire une fouille de processus rapide, facile et tout simplement amusant ;
- ARIS Express : a été initialement développé par IDS Scheer, acheté par Software AG en décembre 2010. C'est un outil de modélisation gratuit pour l'analyse et la gestion des processus métiers. Il prend en charge différentes notations de modélisation telles que BPMN 2, Chaînes de processus événementielles (EPC) ...

6. Découvert de processus : Algorithme α

En process mining, la découverte de processus est l'une des opérations clés qui prend en input les traces du journal des transactions et produit les modèles de processus métier en output. Dans le e-commerce, commander c'est un processus qui commence par une connexion à la page et termine par une livraison de la marchandise commandée. L'application de α -Algorithme, conduit à l'exploration de traces laissés par chaque client ayant réussi ou non à passer sa commande, ayant fait aboutir ou non la procédure de passation de commande pour y ressortir des modèles de processus dans les perspectives :

- D'optimiser le processus ;
- Améliorer la qualité de service ;
- Rendre le processus plus aisé et utilisable (contraintes globales)

En process mining, l'algorithme α est l'un des algorithmes les plus utilisés de l'exploration des données venant du journal de transaction et en déduit des modèles représentant l'exécution réelle du processus par les utilisateurs.

Dans un système d'information, nous avons deux types de données :

- Le point des données : qui découlent de l'exécution complète d'un processus et on les retrouve dans les tables d'une base de données ;
- Les données des procédés : qui sont une constitution factuelle d'activités constituant un processus avec des points d'arrêt ou étapes. on les retrouve dans le fichier journal (event log).

Illustration :

Procédés	Points de données
Consultation	Malaria, Typhoïde, Fièvre...
Transfert en banque	Montant, Date, solde
Achat d'un produit en ligne	Nom du produit, Quantité acheté, date...

Dans son exécution, l'algorithme α présume que les journaux d'événements sont complets ne contienne aucune donnée manquante ou bruit. Dans le déroulement d'un processus qui est une succession d'activités, la détermination des contraintes relationnelles entre pair d'activités demeure une préoccupation. Selon Earth and Environmental Science nous avons 4 contraintes :²

- Le suivi : noté par $a > b$ (a est directement suivi de b)

² Business process mining from e-commerce event web logs: Conformance checking and bottleneck identification, page 3.

International Journal of Novel Research in Computer Science and Software Engineering

Vol. 11, Issue 3, pp: (29-40), Month: September - December 2024, Available at: www.noveltyjournals.com

- Le séquençement : noté par **a b** (si et seulement si a est suivi de b et b ne peut pas être suivi de a) ;
- Le parallélisme : noté par **a # b** (si seulement si a n'est pas suivi de b et b non plus n'est pas suivi de a) ;
- Simultanéité : noté par **a || b** (s'il y a une trace où b suit a et une autre trace où a suit b)

Soit le journal de transaction K constitué de dix traces utilisateurs dont les activités se succèdent de la manière suivante :

- Choix du produit (a) ;
- Mise en panier (b) ;
- Validation panier (c) ;
- Choix mode de paiement (d) ;
- Validation paiement (e) ;
- Confirmation mot de passe (f) ;
- Affichage Adresse de livraison (g) ;
- Confirmation Adresse de Livraison de la commande (h) ;

Partant de la succession normale des activités telles que présentées ci haut, nous pouvons simuler de manière aléatoire et représentative (échantillonnage) un fichier log dont les traces se présenteront de la manière suivante :

- <a, b, c,> ;
- <a, c, d, e> ;
- <a, c, d, >
- <a, b, c, d, e, f, g, h> ;
- <a, b, c, d, e, f, g> ;
- <a, b, c, d, e, f> ;
- <a, c, d, e, f, g, h>

Partant du journal des transactions ci-haut représenté, nous pouvons dégager un certain nombre de profils tels processus exécuté par chacun des clients ayant visité le site marchand. Ici, nous sommes appelé à classifier les clients selon leur habitudes communes dans l'utilisation de la plateforme.

Concernant notre journal de transactions, nous constatons qu'il y a ceux qui s'arrêtent au choix du mode de paiement pour faute d'avoir le moyen de paiement adéquat pour ce faire, à la validation du panier, à la confirmation du mot de passe, et ceux aussi qui termine leur processus en confirmant l'adresse de livraison de la commande.

7. Le degré de complétude

Le degré de complétude est une mesure de la quantité d'informations contenues dans un ensemble de données. Il est généralement exprimé en pourcentage et est calculé en divisant le nombre d'éléments complets par le nombre total d'éléments.

Par exemple, si un ensemble de données contient 100 éléments et que 90 d'entre eux sont complets, le degré de complétude est de 90 %.

Le degré de complétude est important dans de nombreux contextes, tels que les enquêtes, les études de marché et les bases de données. Il permet d'évaluer la qualité des données et d'identifier les éléments qui doivent être complétés.

Il existe plusieurs méthodes pour calculer le degré de complétude. La méthode la plus simple consiste à utiliser la formule suivante :

$$\text{Degré de complétude} = (\text{Nombre d'activités réalisées} / \text{Nombre d'activités attendues}) * 100$$

Dans cette formule, NAR est le nombre d'activités réalisées et NAA est le nombre d'activités attendues.

Dans certains cas, il peut être nécessaire d'utiliser une méthode plus complexe pour calculer le degré de complétude. Par exemple, si certaines activités peuvent être considérées comme plus importantes que d'autres, il peut être nécessaire de pondérer les activités en fonction de leur importance, mais dans notre cas toutes les activités se valent.

Le degré de complétude est un outil précieux pour évaluer la qualité des données. Il permet d'identifier les éléments qui doivent être complétés et de garantir que les données sont suffisamment complètes pour être utilisées à des fins d'analyse.

Voici quelques exemples d'utilisation du degré de complétude :

- Enquêtes : Le degré de complétude peut être utilisé pour évaluer la qualité des données collectées lors d'une enquête. Si le degré de complétude est faible, cela peut indiquer que les participants n'ont pas répondu à toutes les questions ou qu'ils ont répondu de manière incomplète ;
- Etudes de marché : Le degré de complétude peut être utilisé pour évaluer la qualité des données collectées lors d'une étude de marché. Si le degré de complétude est faible, cela peut indiquer que les répondants n'ont pas fourni toutes les informations nécessaires,
- Bases de données : Le degré de complétude peut être utilisé pour évaluer la qualité des données contenues dans une base de données. Si le degré de complétude est faible, cela peut indiquer que la base de données est incomplète ou inexacte.

En conclusion, le degré de complétude est une mesure importante de la qualité des données. Il permet d'identifier les éléments qui doivent être complétés et de garantir que les données sont suffisamment complètes pour être utilisées à des fins d'analyse.

Dans manière pratique, partant de notre journal de transactions ci-haut représenté, pour chaque trace découverte dans notre fichier journal, nous pouvons calculer son degré de complétude afin d'arriver à comprendre pourquoi pour la plupart de cas les clients s'arrêtent telle ou telle autre activité ? Et prendre des dispositions d'amélioration du processus en adaptant ce dernier au vouloir des clients.

Etant donné que le NAA est égal à 8, à chaque trace nous prendrons le NAR divisé par 8 multiplier par 100.

$$DC = \frac{NAR}{NAA} \times 100$$

Pour la première trace le DC est de 37,5%, la deuxième trace 50%, la troisième 37,5%, quatrième 100%, cinquième 87,5% et la sixième, la DC est égal à 75% et enfin pour la septième trace 87,5%.

Ceci nous permet de faire une projection sur la possibilité qu'un client arrive à achever le processus en se faisant livrer les produits après avoir conclu avec succès l'opération en prenant en compte les indicateurs classiques pour déterminer le profil d'un client :

- Le degré de complétude ;
- Le temps d'exécution d'une transaction.

8. Modélisation des données du journal

Dans un système d'information qui a pour rôle d'acquérir, stocker, traiter les données et restituer le résultat, l'acquisition de ces données et la restitution des résultats sont tributaires d'un ensemble des tâches qui constituent un processus. L'exécution de ces deux dernières tâches dans le système d'information laisse des traces dans le journal d'événements qui constituent une base de données spécifique dans laquelle tous les événements ou processus sont stockés avec horodatage informations comme signatures (empreintes). Chaque trace est constituée des trois principales valeurs qui sont :

- Le numéro d'identification du cas (Id) ;
- L'activité exécutée ;
- L'horodatage
- L'auteur

Exemple :

Case ID	Activity	Time Stamp	User	Cost
1	Incident logging	2022/02/20 12:09:00	edmund.price	20.00
1	Incident classification	2022/02/20 12:10:20	edmund.price	20.00
2	Incident logging	2022/02/20 12:33:01	angela.lewis	20.00
1	Initial diagnosis	2022/02/20 12:33:03	philip.sanchez	20.00
2	Incident classification	2022/02/20 12:34:11	angela.lewis	40.00
2	Initial diagnosis	2022/02/20 12:41:08	chris.watkins	30.00
1	Functional escalation	2022/02/20 15:50:27	christine.chenal	10.00
1	Investigation and diagnosis	2022/02/20 9:24:43	christine.chenal	10.00
1	Resolution and recovery	2022/02/20 10:44:01	christine.chenal	10.00
1	Incident closure	2022/02/20 11:57:21	karen.anderson	10.00

En outre, toutes les autres valeurs de données dans le journal d'événements sont considérées comme des données supplémentaires (c'est-à-dire la valeur de la transaction, personne ayant déclenché la transaction). Une entreprise dans son fonctionnement a dans la plupart de cas un ensemble de processus métier qui pourrait constituer un certain nombre de cycles de processus ou de sous-processus pour un but bien précis.

Un cycle de processus de sous-processus est constitué d'un scénario dans les données du journal des événements, par ex. un scénario de cycle de traitement de commande ou de sous-processus de retour d'article. Un certain nombre d'entreprises ne disposent probablement pas de journaux d'événements de leurs processus métier ; cependant, ils avaient l'habitude d'avoir des enregistrements horodatés de certaines activités³. Les horodatages des activités peuvent être obtenus, par exemple, à partir de l'événement de vente organisé, du paiement effectué, du produit livré, etc.

9. Présentation du Journal

✚ Environnement de présentation

Dans le cadre de notre étude, nous avons utilisé Flexicon Disco qui est un logiciel de process mining développé par la société Fluxico. Disco est utilisé pour analyser et visualiser les processus métier afin d'identifier les inefficacités, les goulots d'étranglement et les domaines à améliorer.

Il peut être également utilisé pour analyser tout type de processus, y compris les processus orientés client, les processus internes et les processus de fabrication. Aussi pour analyser les processus répartis sur plusieurs systèmes et départements. Disco utilise diverses techniques pour analyser les processus, notamment :

- Analyse des journaux d'événements : Flexicon Disco peut analyser les journaux d'événements de tout type de système, y compris les systèmes ERP, les systèmes CRM et les systèmes d'exécution de fabrication (MES).
- Cartographie des processus : Flexicon Disco peut créer des cartes de processus qui montrent le flux des activités dans un processus.
- Exploration de données : Flexicon Disco peut utiliser des techniques d'exploration de données pour identifier des modèles et des tendances dans les données de processus.

Flexicon Disco fournit une variété de rapports et de tableaux de bord qui peuvent être utilisés pour visualiser et analyser les données de processus. Ces rapports et tableaux de bord peuvent être utilisés pour identifier les goulots d'étranglement, les inefficacités et les domaines à améliorer.

✚ Présentation du fichier journal avec Disco

Nous avons simulé un fichier journal de manière aléatoire partant sur une présentation ou succession d'activités dans un processus de vente en ligne. Notre journal de transaction est structuré en 4 rubriques dont :

- Cas :

³ International Conference on Biospheric Harmony Advanced Research 2020 : **Business process mining from e-commerce event web logs: Conformance checking and bottleneck identification**

- Activités :
- Horodatege :
- Identifiant :

Spécification de l'algorithme

Le journal de transactions sera généré sur base d'un algorithme spécifique tel que présenté ci-dessous :

1. ****Initialiser les variables :****

- `roll_no` ← 9
- `v` ← Entier (utilisé pour stocker des valeurs aléatoires)
- `caseID`, `Activity`, `Etimes`, `custID` ← Chaînes de caractères
- `listActivities` ← Tableau de 8 chaînes de caractères contenant différentes activités
- `fout` ← Objet de flux de fichier pour la sortie
- `mois`, `jour`, `heure`, `minutes`, `secondes` ← Entiers pour stocker les composants de la date et de l'heure

2. ****Ouvrir/Créer un fichier CSV :****

- Ouvrir le fichier `ECommerce_Log.csv` en mode écriture avec `fout`.
- Écrire l'en-tête `"Case No,Activity,Timestamp,CustomerID"` dans le fichier.

3. ****Générer des entrées de journal :****

- ****Pour**** `i` de 1 à 91 :
 - ****Pour**** `j` de `i` à 91 :
 1. Affecter `"Case"` à `caseID`.
 2. Afficher `"Pour Case : Case"` et la valeur de `i`.
 3. Générer une valeur aléatoire `v` entre 0 et `roll_no-1`.
 4. Afficher `"random value"` et la valeur de `v`.
 5. Affecter à `Activity` l'élément `v` du tableau `listActivities`.
 6. Afficher `"Activity :"` et la valeur de `Activity`.
 7. Générer les composants du timestamp :
 - `mois` ← `i + 1`
 - `jour` ← Valeur aléatoire entre 0 et 29
 - `heure` ← Valeur aléatoire entre 0 et 23
 - `minutes` ← Valeur aléatoire entre 0 et 59
 - `secondes` ← Valeur aléatoire entre 0 et 59
 8. Affecter la chaîne `"2022"` à `Etimes`.
 9. Afficher `"Timestamp :"` et la valeur de `Etimes`.
 10. Affecter `"Client"` à `custID`.
 11. Afficher `"Client "` et la valeur de `custID`.
 12. Écrire les données suivantes dans le fichier CSV :
 - `"Case ", i, Activity, Etimes, mois, jour, heure, minutes, secondes, custID, i`

4. ****Fermer le fichier CSV :****

- Fermer le fichier `fout`.

5. ****Afficher le contenu du fichier CSV :****

- Ouvrir le fichier `ECommerce_Log.csv` en mode lecture avec `fin`.

- Afficher `\"CSV file has been successfully created..!\"`.

- Afficher `\"Displaying the content of CSV file=>\n\n\"`.

- ****Pour**** `i` de 0 à 500 :

- Lire une ligne du fichier CSV et la stocker dans `data`.

- Afficher le contenu de `data`.

6. ****Fin du programme :****

- Retourner 0 pour indiquer une exécution réussie.

Implémentation de l'algorithme (Codes sources)

```
#include<iostream>
```

```
#include<fstream> //used for file handling
```

```
#include<string> //used for strings
```

```
#include<cstdlib>
```

```
using namespace std;
```

```
int main()
```

```
{
```

```
    int roll_no = 9, v;
```

```
    string caseID,Activity,etimes,custID;
```

```
    string listActivities [8] = {"Choix du produit (a)","Mise en panier (b)","Validation panier (c)",
```

```
                                "Choix mode de paiement (d)","Validation
```

```
paiement (e)","Confirmation mot de passe (f)",
```

```
                                "Affichage Adresse de livraison
```

```
(g)","Confirmation Adresse de Livraison de la commande (h)"};
```

```
    fstream fout;
```

```
    // opening an existing csv file or creating a new csv file
```

```
    fout.open("ECommerce_Log.csv", ios::out);
```

```
    fout<<"Case No"<<" "<<"Activity"<<" "<<"Timestamp"<<" "<<"CustomerID"<<"\n";
```

```
    int mois,jour,heure,minutes,secondes;
```

```
    for(int i=1; i <= 91; i++)
```



```

{
    for(int j=i; j <= 91; j++)
        {
            caseID = "Case";
            cout<<"Pour Case :" << caseID<<" "<<i<<"\n";
            v = rand() % roll_no;
            cout<<"random value "<<v<<"\n";
            Activity = listActivities[v];
            cout<<"Activity : " <<Activity<<"\n";
            // Etimes = "2022"+ i + rand() % 30 + " " + rand() % 18 + " " + rand() % 60 + " " +
            rand() % 60;

            mois = i+1;
            jour = rand() % 30;
            heure = rand() % 24;
            minutes = rand() % 60;
            secondes = rand() % 60;

            Etimes = "2022";
            cout<<"Timestamp :" <<Etimes<<"\n";
            custID = "Client";
            cout<<"\nClient " <<custID<<"\n";

            fout<<caseID<<" "<<i<<","<<Activity<<","<<Etimes<<"/"<<mois<<"/"<<jour<<"
            "<<heure<<":"<<minutes<<":"<<secondes<<","<<custID<<" "<<i<<"\n";

        }
}

fout.close(); // closing csv file

fstream fin;
string data;
cout<<"\nCSV file has been successfully created..!!";
cout<<"\n\nDisplaying the content of CSV file=>\n\n";
    // opening existing csv file in read mode
fin.open("ECommerce_Log.csv", ios::in);

```

```
for(int i=0; i<=500; i++)
{
    // getline fun will read one entire row of csv file and will store it in string 'data'
    getline(fin,data);
    cout<<data<<endl;
}

return 0;
}
```

Journal de transactions généré

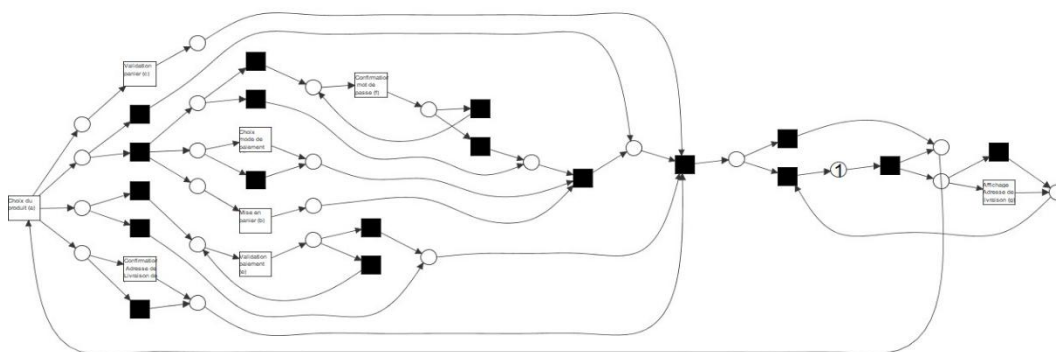
event	Activity	Timestamp	CustomerID
1	Case 1 Confirmation mot de passe (f)	2022/2/17 22:40:29	Client 1
2	Case 1 Mise en panier (b)	2022/2/18 6:22:44	Client 1
3	Case 1 Client	2022/2/5 1:27:1	Client 1
4	Case 1 Confirmation mot de passe (f)	2022/2/25 14:27:36	Client 1
5	Case 1 Choix du produit (a)	2022/2/24 14:33:52	Client 1
6	Case 1 Confirmation Adresse de Livraison de la command...	2022/2/21 20:38:35	Client 1
7	Case 1 Validation panier (c)	2022/2/6 11:18:9	Client 1
8	Case 1 Validation paiement (e)	2022/2/17 19:55:54	Client 1
9	Case 1 Validation panier (c)	2022/2/21 2:33:33	Client 1
10	Case 1 Validation panier (c)	2022/2/21 7:53:28	Client 1
11	Case 1 Confirmation mot de passe (f)	2022/2/14 22:57:57	Client 1
12	Case 1 Confirmation Adresse de Livraison de la command...	2022/2/23 21:49:58	Client 1
13	Case 1 Validation paiement (e)	2022/2/5 14:42:48	Client 1
14	Case 1 Mise en panier (b)	2022/2/19 14:4:28	Client 1
15	Case 1 Confirmation mot de passe (f)	2022/2/15 2:9:10	Client 1
16	Case 1 Confirmation mot de passe (f)	2022/2/6 21:33:8	Client 1
17	Case 1 Choix du produit (a)	2022/2/23 12:34:36	Client 1
18	Case 1 Confirmation mot de passe (f)	2022/2/16 8:11:28	Client 1
19	Case 1 Affichage Adresse de livraison (g)	2022/2/9 2:43:17	Client 1
20	Case 1 Mise en panier (b)	2022/2/8 18:9:41	Client 1
21	Case 1 Choix du produit (a)	2022/2/5 7:18:24	Client 1
22	Case 1 Choix mode de paiement (d)	2022/2/27 2:53:6	Client 1
23	Case 1 Client	2022/2/6 20:52:30	Client 1
24	Case 1 Affichage Adresse de livraison (g)	2022/2/17 21:57:12	Client 1
25	Case 1 Mise en panier (b)	2022/2/0 17:36:35	Client 1
26	Case 1 Client	2022/2/15 22:11:52	Client 1
27	Case 1 Client	2022/2/10 21:44:45	Client 1
28	Case 1 Mise en panier (b)	2022/2/27 23:7:57	Client 1
29	Case 1 Validation paiement (e)	2022/2/17 9:3:6	Client 1

Fig. Journal de Transaction

Présentation du modèle avec PROM

Le modèle étant une représentation schématique d’une réalité tout en respectant certaines règles et principes, le nôtre présente de manière succincte toutes les activités réalisées par les clients sur le site marchand HP sous forme d’un réseau de pétri.

Statistiques d’exécution des tâches



Log Summary		
Total number of process instances: 1		
Total number of events: 76		
Event Name		
Event classes defined by Event Name		
All events		
Total number of classes: 8		
Class	Occurrences (absolute)	Occurrences (relative)
Confirmation mot de passe (f)	15	19,737%
Validation panier (c)	14	18,421%
Validation paiement (e)	12	15,789%
Confirmation Adresse de Livraison de la commande (h)	11	14,474%
Choix du produit (a)	8	10,526%
Mise en panier (b)	8	10,526%
Choix mode de paiement (d)	5	6,579%
Affichage Adresse de livraison (g)	3	3,947%
Start events		
Total number of classes: 1		
Class	Occurrences (absolute)	Occurrences (relative)
Choix du produit (a)	1	100,0%
End events		
Total number of classes: 1		
Class	Occurrences (absolute)	Occurrences (relative)
Validation paiement (e)	1	100,0%
(Event Name AND Lifecycle transition)		
Event classes defined by (Event Name AND Lifecycle transition)		
All events		
Total number of classes: 8		
Class	Occurrences (absolute)	Occurrences (relative)
Confirmation mot de passe (f)+start	15	19,737%
Validation panier (c)+start	14	18,421%
Validation paiement (e)+start	12	15,789%
Confirmation Adresse de Livraison de la commande (h)+start	11	14,474%
Choix du produit (a)+start	8	10,526%
Mise en panier (b)+start	8	10,526%
Choix mode de paiement (d)+start	5	6,579%
Affichage Adresse de livraison (g)+start	3	3,947%
Start events		
Total number of classes: 1		
Class	Occurrences (absolute)	Occurrences (relative)
Choix du produit (a)+start	1	100,0%
End events		
Total number of classes: 1		
Class	Occurrences (absolute)	Occurrences (relative)
Validation paiement (e)+start	1	100,0%

10. Profils identifiables

Partant de notre graphe ou modèle tel que présenté à la page précédente, en considérant les deux indicateurs classiques dans la détermination du profil client, nous avons dégagé cinq (5) profils dont :

- Ceux qui s'arrêtent à la validation du panier avec 18,42% de DC ;
- Ceux qui s'arrêtent à la confirmation mot de passe 19,73% de DC ;
- Ceux qui s'arrêtent à la confirmation du paiement 15,78% de DC ;
- Ceux qui s'arrêtent à la confirmation de l'adresse de livraison 3,94% de DC ;
- Ceux qui s'arrêtent au choix du mode de paiement 6,57%

2. CONCLUSION

Dans le cadre de l'analyse des journaux de transactions du système d'information, la mise en place d'une stratégie d'exploration et d'analyse des journaux de transactions présente de nombreux bénéfices dans l'identification des écarts, améliorer les processus et proposer des biens et services adaptés aux besoins des utilisateurs, tout en favorisant leur fidélisation.

Une stratégie d'exploration et d'analyse des journaux de transactions permet, dans un premier temps, d'identifier les écarts entre les processus prévus et leur exécution réelle. Les journaux contiennent des traces détaillées des activités effectuées par les utilisateurs dans le système d'information. En analysant ces journaux, il est possible de détecter les moments où les utilisateurs s'écartent des procédures établies, ce qui peut indiquer des problèmes potentiels ou des inefficacités dans le processus. Par exemple, si un utilisateur suit un parcours différent pour remplir une demande en ligne, cela peut être considéré comme un écart et nécessiter une enquête pour comprendre les raisons sous-jacentes.

L'identification de ces écarts permet d'améliorer les processus existants en les ajustant pour mieux répondre aux besoins réels des utilisateurs. En examinant les données des journaux, il est possible d'identifier les points de congestion, les goulets d'étranglement ou les confusions rencontrées par les utilisateurs dans leur parcours. Grâce à cette analyse, les responsables peuvent profiler les clients selon leurs habitudes dans l'utilisation de la plateforme partant de deux indicateurs tels que la durée d'exécution et le degré de complétude tel que proposé dans cet article.

Une fois que les écarts sont identifiés, les processus sont améliorés et profilés déterminés, une stratégie d'exploration et d'analyse des journaux de transaction permettra également de proposer des biens et services adaptés aux besoins des utilisateurs. En analysant les traces laissées par les utilisateurs dans les journaux, il est possible d'extraire des informations précieuses sur leurs préférences, leurs intérêts et leurs comportements. Ces informations peuvent être utilisées pour personnaliser les offres proposées aux utilisateurs, en leur recommandant des produits ou des services qui correspondent à leurs besoins spécifiques. Par exemple, si un utilisateur consulte fréquemment des articles sur la décoration intérieure, une stratégie d'analyse des logs pourrait lui recommander des produits liés à ce domaine lorsqu'il se connecte au système.

BIBLIOGRAPHIE

- [1] M Ghasemi, D Amyot (2016). Process mining in healthcare ;
- [2] W Van Der Aalst, A Adriansyah, W Van Der Aalst, V De Weerd (2012). Process mining manifesto. Business Process Management ;
- [3] O Benreskallah (2014). Une approche de Data Mining Visuel.
- [4] D Peraya, C Batier, D Paquelin (2009). Les traces d'usage et l'usage des traces;
- [5] L Yahiaoui, Y Prié, Z Boufaïda (2008). Redocumentation des traces d'activité médiée informatiquement dans le cadre des transactions communicationnelles.
- [6] M Assaf (2022). Le rôle du marketing relationnel dans la fidélisation des clients Cas: ENIEM.