

## Data Mining au service du marketing digital: Cas des produits bancaires

### [ Data Mining at the service of digital marketing: Case of banking products ]

*Amzile Rajaa and Amzile Karim*

Département sciences de gestion, Université Mohammed V Rabat, Faculté des sciences juridiques, économiques et sociales,  
Agdal, Rabat, Morocco

Copyright © 2020 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** Through this literature review, we have explored the various scientific articles that have been exploited by Data Mining techniques in the banking sector and specifically Digital Marketing, namely the prediction of customers interested in a banking product. We have succeeded through a variety of articles in tracking the behavior of some Data Mining algorithms and their uses, using bank data. These techniques have also allowed us to determine the best parameters for each Data Mining method. In this review, we have used the results of different evaluation tools for all the methods used, which has enabled us to easily choose the appropriate prediction model. Several Software have been used as a programming tool to clarify each of the techniques of Data Mining, Therefore, we cannot say that an algorithm is the best since the results differ from one category to another of the data used.

**KEYWORDS:** Big data, data mining, digital, marketing, Genetic Algrothm.

**RESUME:** À travers cette revue de littérature, nous avons exploré les différents articles scientifiques qu'ont exploité les techniques du Data Mining dans le secteur bancaire et bien précisément le Marketing digital à savoir la prédiction des clients intéressés par un produit bancaire. Nous avons réussi à travers une panoplie d'articles de suivre le comportement de quelques algorithmes du Data Mining ainsi que leurs utilités, en utilisant des données bancaires. Ces techniques nous ont également permis de déterminer les meilleurs paramètres de chaque méthode du Data Mining. Dans cette revue, nous avons exploités les résultats de différents outils d'évaluation de toutes les méthodes utilisées, ce qui nous a permet de choisir avec aisance le modèle de prédiction adéquat. Plusieurs Logiciels ont été utilisé comme outil de programmation pour éclaircir chacune des techniques du Data Mining, Par conséquent, nous ne pouvons pas dire qu'un algorithme est le meilleur vu que les résultats se différent d'une catégorie à une autre de données utilisées.

**MOTS-CLEFS:** Big data, data mining, digital, marketing, algorithme génétique.

## 1 INTRODUCTION

À l'ère des nouvelles technologies le marketing est devenu l'un des aspects qui ont su exploiter ce monde digital, et parmi les techniques qui ont su exploiter il y a celles du data ming qu'ont pour rôle d'explorer un grand volume de données, on lui applique une panoplie de techniques et méthodes. Le Datamining<sup>1</sup> peut aider les commerciaux pour arriver au client cible avec aisance. Le secteur bancaire a soif de ce genre d'outils qui vont lui permettre d'optimiser son marché de produits financiers, alors que c'est le secteur qu'est en plein d'expansion en matière de données stockées.

---

<sup>1</sup> Ensemble de techniques et des outils qui nous aide à faire des extractions de connaissances à partir de données.

Aujourd'hui, il est devenu primordial de se focaliser sur sa clientèle afin d'être plus concurrentiel. Avoir un client sûr est la recette pour épanouir ses revenus et élargir son chiffre, les techniques du Data Mining vont aider les banques à réaliser plus de succès dans la commercialisation d'un produit particulier à travers des données bien choisies et traitées.

## 2 MATERIEL ET METHODES

### 2.1 UNE VUE GENERALE SUR LES ARTICLES ETUDIES

A travers cette revue de littérature on a veillé à faire une lecture approfondie et pertinente a des articles scientifiques de valeur, on a essayé de choisir un intervalle de temps cohérent qui représente l'année de l'édition, allant de l'année 2002 jusqu'à l'année 2020. Et à travers ce tableau ci-dessous vous allez trouver les différentes clés qui caractérisent chaque un des articles revus.

Tableau 1. Revue d'articles scientifiques pour la période 2002-2020

	Année	Auteurs	Mots Clé	Journal
The Implementation of Subspace Outlier Detection in K-Nearest Neighbors to Improve Accuracy in Bank Marketing Data <sup>1</sup>	2020	Dimas Aryo Anggoro <sup>1</sup> PuteralIslamiyadi Rahmatullah <sup>2</sup>	Direct Marketing, KNN, PCA, SOD <sup>1</sup> .	International Journal of Emerging Trends in Engineering Research
Visualization and Analysis in Bank Direct Marketing Prediction <sup>2</sup>	2019	Alaa Abu-Srhan Bara'aAlhammad Sanaa Al zghoul Rizik Al-Sayyed	Direct Marketing, Prediction, Visualization, Oversampling, Naïve Bayes.	International Journal of Advanced Computer Science and Applications, Vol. 10, No. 7, 2019
Term Deposit Subscription Prediction Using Spark MLlib and ML Packages <sup>3</sup>	2019	Phan Duy Hung Tran Duc Hanh Ta Duc Tung	Bank Marketing; Predicting Term Deposit Subscriptions; Apache Spark; MLlib; ML packages; Decision Tree (DT); Random forest (RF); Gradient-boosted trees (GBT).	ICEBA 2019, February 25–28, 2019, BANGKOK, Thailand
Predicting customer response to bank direct telemarketing campaign <sup>4</sup>	2017	Justice Asare-Frempong ManojJayabalan	Bank, telemarketing, Machine learning Techniques, Clustering, Direct marketing	International Conference on Engineering Technology and Technopreneurship (ICE2T)
A Data Mining Approach for Bank Telemarketing Using the rminer Package and R Tool <sup>5</sup>	2013	Sérgio Moro Paulo Cortez Raul M. S. Laureano	Telemarketing, Direct Marketing, Long-term Deposits, Data Mining, CRISP-DM, Classification Problem, Banking, R.	ISCTE – Instituto Universitário de Lisboa BRU-IUL (Unide-IUL)
An intelligent system for customer targeting a data mining approach <sup>6</sup>	2002	YongSeog Kim W. Nick Street	Customer targeting; Data mining; Feature Selection; Genetic Algorithms; Neural Networks; Ensemble	Decision Support Systems

### 2.2 LA PROBLEMATIQUE TRAITEE

Depuis l'arrivée des nouvelles technologies dans divers domaines et en présence d'une évolution progressive d'internet, le marketing, à l'instar des autres domaines, s'est aussi mis au niveau de ces évolutions technologiques qui l'ont rendu digital d'où la nouvelle dénomination « marketing digital » apparue en 1992 avec la naissance des moteurs de recherches sur internet. Ce nouveau concept a donné une nouvelle définition au marketing en exploitant des données numériques à travers des outils informatiques dédiés à analyser et produire des informations utiles afin de créer des stratégies assez robustes et fiables. A cet effet, la question qui se pose: comment les techniques du Data Mining peuvent servir le marketing digital ?

Afin de répondre à cette question il y a lieu de Comprendre:

Qu'est-ce que le Marketing traditionnel et comment il a muté vers la digitalisation ? Quels sont les principaux acteurs du marketing digital ?

Comment réussir sa stratégie marketing digital ?

### 2.3 DEFINITION GENERALE DU MARKETING

Il n'existe pas de formule universelle pour l'organisation du travail et pour la mobilisation des moyens dans la sphère de marketing. Encore plus, il n'existe pas de schéma unique de construction du système marketing. Bref, il n'y a pas d'unité de vue sur le concept marketing:

- Selon Peter Drucker: Le but du marketing est de rendre la vente superflue. Une entreprise n'a que deux fonctions essentielles: le marketing et l'innovation, qui produisent des résultats, tout le reste ne produit que des coûts;
- Selon l'Association américaine de marketing: Le marketing est la réalisation de différents types d'activités économiques qui orientent et conduisent les flux de marchandises et de services depuis le producteur jusqu'à l'utilisateur final ou intermédiaire;
- Selon Philippe Kotler: Le marketing est le mécanisme économique et social par lequel l'individu et groupes satisfont leurs besoins et désirs au moyen de la création et de l'échange de produits et d'autres entités de valeur pour autrui.

Les auteurs des différentes définitions du marketing accordent une attention particulière à deux éléments essentiels:

- Le marketing en tant que système
- Le marketing comme méthode

Ainsi, Le marketing forme un système complexe d'organisation de la production et de la vente des produits basées sur la recherche et l'étude préliminaire des besoins des consommateurs.

La conception du marketing s'étend non seulement à l'organisation et à la réalisation des produits déjà fabriqués, mais avant tout à l'étude du marché pour l'orientation ultérieure vers la production de marchandises vendables. De plus, le marketing doit contribuer à l'accroissement des bénéfices de l'entreprise tout en essayant de répondre aux exigences du consommateur.

### 2.4 L'ESSENCE MARKETING

A l'instar des autres sciences, le marketing a également une origine et une histoire. Toutefois, le marketing est considéré comme discipline plutôt jeune.

On peut diviser l'histoire du marketing en 5 principales phases:

#### • Phase 1: L'économie de subsistance

Le concept de marketing couvrait les activités réalisées par les gens qui cherchaient à négocier des conditions d'échanges mutuellement convenables.

Ainsi, Les activités d'échange encouragèrent les gens à apporter différentes quantités de marchandises en un lieu commun, ce qui donna naissance aux marchés locaux. C'est à ce moment qu'on assiste à la naissance des premiers vendeurs, puis l'utilisation d'un moyen d'échange commun: la monnaie qui a permis d'accélérer le rythme du business.

Le marketing se synthétisait alors aux: activités visant l'échange de biens économiques et donc il suffisait de produire pour vendre.

#### • Phase 2: la révolution industrielle

Avec l'accélération du trafic, les gens s'enfoncèrent à échanger pour le profit.

D'énormes ressources furent condensées dans des industries, la révolution industrielle avait lieu. Cette ère se caractérise par la pénurie des moyens de production.

A cette époque, les besoins à satisfaire étaient les besoins de base, il fallait organiser une production de masse afin de répondre à la demande du marché, et ce au coût de revient, éventuellement le plus faible.

Ainsi, Le rôle du marketing se délimitait à prospecter des marchés sur lesquels écouler les produits fabriqués. C'était un marketing passif: il n'y avait aucune prise en considération des besoins du consommateur.

- **Phase 3: La croissance économique**

Cette phase se détermine par une production de quantité importante et en conséquence, par un certain détroit dans les circuits de distributions traditionnels.

Les marchés étaient géographiquement de plus en plus éloignés des producteurs ainsi les circuits de distribution se prolongeaient. A cet effet, de nouvelles formes de communication ont été innovées: la publicité de masse est née. Ainsi, le rôle de marketing consistait à trouver des débouchés afin d'améliorer et de stimuler l'économie des biens fabriqués: entreprise orientée vers la vente où le marketing joue un rôle d'organisateur et de promoteur (on incite l'acheteur à consommer ce qu'on produit).

- **Phase 4: L'économie d'abondance**

Les objectifs de la croissance économique sont atteints, on est maintenant présent face à une situation de rareté de la demande. Le marché global n'étant plus suffisant, on s'est focalisé à des segments plus réduits de ce marché. Comme leurs besoins primordiaux étaient satisfaits, les consommateurs pouvaient se diriger vers des besoins plus psychologiques (besoins de sécurité, d'épanouissement...).

L'entreprise a, en conséquence, essayé de mettre à niveau sa capacité de production ainsi que ses lignes de produits aux désirs des consommateurs.

Ainsi, le rôle du marketing est passé de passif à très actif: l'objectif du marketing a alors été d'adapter les produits et les services offerts aux besoins prévus par le marché.

- **Phase 5: La société post industrielle**

L'entreprise doit maintenant tenir compte des retombées technologique, sociologiques et écologiques. Le rôle du marketing est maintenant de créer de la valeur pour chaque client et le fidéliser. L'apparition des nouvelles technologies d'informations et de communication a inversé les choses, Internet se démocratise et change complètement les moyens de consommation, et par conséquent les moyens de communication: C'est la naissance du marketing sur le web ou encore le marketing digital.

## 2.5 LE MARKETING DIGITAL.

Le marketing digital désigne l'ensemble des techniques Marketing qui utilisent les tunnels digitaux et l'exploration des données de masse (big data<sup>2</sup>). Il se base surtout sur les applications marketing liées à Internet, mais également celles liées aux smart phones, tablettes et autres applications. Le marketing digital se distingue du marketing traditionnel par de nombreux aspects qui introduisent une réelle rupture et un nouveau paradigme. Il va au-delà du e-marketing<sup>3</sup> puisqu'il couvre l'ensemble des médias digitaux, tout en cherchant à fonder des relations plus directes et ciblées avec les consommateurs. Parmi les particularités ou éléments distinctifs du marketing digital, on peut citer entre autres un marketing qui s'effectue en temps réel, qui utilise des données de grande masse, un marketing souvent facturé à la performance, un marketing mathématique et algorithmique<sup>4</sup>.

Le marketing digital suppose de gérer la présence de l'organisation ou du produit dans le monde digital, dans les médias et les plateformes en ligne<sup>5</sup>. Ces diverses technologies permettent d'atteindre des objectifs de recrutement de nouveaux clients, mais aussi de fidéliser la clientèle actuelle, de développer l'e-réputation<sup>6</sup> d'une entreprise ou d'un produit.

### 2.5.1 LES ELEMENTS CLES D'UNE STRATEGIE DE MARKETING DIGITAL

Les questions de segmentation, de ciblage et de positionnement se posent dans le marketing digital comme dans le marketing classique: elles sont d'autant plus importantes qu'il s'agit de coordonner l'ensemble des actions, digitales ou non,

---

<sup>2</sup> Désigne des ensembles de données devenus si volumineux et massives

<sup>3</sup> Correspond à l'ensemble des méthodes et des pratiques Marketing sur Internet

<sup>4</sup> Source : site ABC Net marketing

<sup>5</sup> Les sites web, moteurs de recherche, réseaux sociaux, applications mobiles, e-mails

<sup>6</sup> C'est la réputation du produit et de l'entreprise dans le monde digital

pour atteindre les objectifs fixés par le marketing. L'intégration entre les canaux, digitaux (online) et traditionnels (offline), est une dimension essentielle de la stratégie digitale. Cette dernière est avant tout une stratégie de canal dans le contexte d'une stratégie cross-canal plus large. Par conséquent, une stratégie digitale doit :

- Préciser l'orientation future des activités de marketing digital et définir des objectifs qui soutiennent les objectifs Marketing et commerciaux de l'entreprise, en cohérence avec sa stratégie d'affaires;
- Décrire comment les ressources seront déployées et comment l'organisation doit se structurer pour mettre en œuvre cette stratégie;
- Prendre en compte les dimensions de toute stratégie marketing, à savoir la définition des marchés cibles, le positionnement et la définition du mix marketing<sup>7</sup>;
- Préciser le mix de communication en ligne et hors ligne pour attirer les cibles sur les médias possédés par l'organisation et les modalités de contact sur l'ensemble des autres médias;
- Accompagner les clients tout au long de leur parcours et de leur cycle de vie (du statut de prospect<sup>8</sup> jusqu'à, idéalement, celui d'ambassadeur de la marque).

## 2.6 LES DIFFERENTES TECHNIQUES ET METHODES DU DATA MINING POUR REUSSIR SA STRATEGIE NUMERIQUE DU MARKETING

### 2.6.1 QU'EST CE QUE LE DATA MINING ?

Le Data Mining, peut se traduire par la prospection ou la fouille de données, connaît un développement commercial très important. Il se base sur des formules mathématiques et statistiques ainsi des algorithmes et des outils informatiques. La motivation principale de cette démarche est la valorisation d'une grande base ou entrepôt de données par la recherche d'informations pertinentes pour l'aide à la décision. Les applications les plus répandues concernent la recherche d'une typologie de la clientèle ou encore celle de scores mesurant l'adéquation des clients aux produits à promouvoir en vue. D'autres applications voient le jour dans un contexte industriel ou encore dans des disciplines scientifiques confrontées à la détection de motifs ou formes dans des volumes de données considérables. Le data Mining peut contribuer à la résolution des problèmes décisionnels dans les domaines de la banque, de la finance par la détection de patterns, de causalités et/ou de corrélations dans les données qui ne sont pas immédiatement décelables par les managers du fait d'un grand volume de données. Les décideurs dans ces domaines peuvent aussi utiliser le data Mining pour identifier des séquences ou des périodicités de certains comportements transactionnels de leurs clients ce qui leur permet d'améliorer la segmentation, le ciblage et la maintenance d'une base de clients profitables.

### 2.6.2 QUELLES SONT LES DIFFERENTES METHODES UTILISEES PAR LE DATA MINING ?

Dans les méthodes utilisées par le Data Mining, on distingue deux grandes familles :

- **Les méthodes descriptives:** permettent d'organiser, de simplifier et d'aider à comprendre l'information à partir des sources de données. (Par exemple: recherche d'associations / recherche de séquences similaires...etc.).
- **Les méthodes prédictives:** visent à expliquer ou prévoir plusieurs phénomènes observables et effectivement mesurés. On cherche à prédire la valeur d'une variable cible à partir des valeurs de prédicteurs. (Par exemple: régression linéaire multiple / réseaux de neurones / arbres de régression...) Autrement dit on cherche à anticiper la valeur de quelque chose (par exemple, si un client risque de ne pas pouvoir rembourser un prêt, c'est la variable cible « Target ») en fonction de ses caractéristiques connues (âge, emploi, salaire... ce sont les prédicteurs), en se basant pour cela sur les données dont on dispose (les anciens clients et les valeurs des prédicteurs et des variables cibles)

Toute une terminologie s'est développée autour du Data Mining et de l'informatique décisionnelle (business intelligence<sup>9</sup>). On parle ainsi de « Knowledge Discovery in Databases » (KDD)<sup>10</sup> qui se décompose en différentes étapes très générales dont certaines sont classiques dans le métier de statisticien :

<sup>7</sup> C'est l'ensemble des décisions et d'actions marketing prises pour assurer le succès d'un bien ou un service marchand.

<sup>8</sup> Un probable futur client, soit une personne correspondant aux critères des acheteurs et avec qui une vente sera conclue.

<sup>9</sup> L'ensemble des outils et méthodes visant à transmettre les informations pertinentes au sein de l'entreprise.

<sup>10</sup> L'extraction de connaissances est le processus de création de connaissances à partir d'informations structurées dans une base de données.

1. Compréhension du domaine d'application: expliciter la connaissance a priori et les buts du commanditaire.
2. Création d'un sous-ensemble cible des données (matrice) à partir de l'entrepôt.
3. Nettoyage: erreurs, données manquantes.
4. Transformation des données: "normalisation", linéarisation, découpage en classes, compression.
5. Explication de l'objectif et de la stratégie d'analyse: exploration, association, classification, discrimination, recherche de formes...
6. Choix des méthodes, des algorithmes, en privilégiant interprétabilité ou prédictibilité. Mise en œuvre des outils informatiques appropriés.
7. Test: sur la base de critères à préciser (qualité d'ajustement, de prévision, simplicité, visualisations graphiques...).
8. Exploitation.
9. Diffusion des résultats pour aider à la prise de décision.

De façon plus précise, le Data Mining concerne l'exécution des étapes 3 à 8 ci-dessus. Il nécessite la mise en œuvre, de méthodes statistiques classiques (graphiques, sondages, composantes principales, correspondances multiples, classification hiérarchique, nuées dynamiques, discriminante, segmentation, régression linéaire, logistique) ou moins classiques (arbres de classification et de régression, modèles graphiques d'indépendance conditionnelle) ou d'intelligence artificielle<sup>11</sup> (perceptron<sup>12</sup> multicouche, réseaux auto-associatif et bayésien, apprentissage et règles d'induction, reconnaissance de formes).

### 3 RESULTATS

#### 3.1 LES METHODOLOGIES ADOPTEES DANS LES ARTICLES ET LES DONNEES UTILISEES

Tableau 2. *Méthodologie & Données utilisées dans les articles revus*

Titre de l'article	Méthodologie & Données
<b>The Implementation of Subspace Outlier Detection in K-Nearest Neighbors to Improve Accuracy in Bank Marketing Data<sup>1</sup></b>	<ul style="list-style-type: none"> <li>• <b>Données:</b> L'ensemble de données utilisé dans cette recherche concernent les activités de marketing de l'institution bancaire portugaise de 2008 à 2010 où l'agence a effectué des appels téléphoniques contacts pour mener des campagnes de marketing pour le potentielles clients à souscrire à des dépôts à terme.</li> <li>• <b>Méthodologie:</b> le processus qui était suivi dans cette étude est scindé en 4 étapes, à savoir la collection des données, le prétraitement des données<sup>13</sup>, le traitement des données<sup>14</sup>, L'évaluation<sup>15</sup> des modèles obtenus.</li> </ul>
<b>Visualization and Analysis in Bank Direct Marketing Prediction<sup>2</sup></b>	<ul style="list-style-type: none"> <li>• <b>Données:</b> L'ensemble de données est obtenu à partir du site de Machine Learning UCI<sup>16</sup>. Cet ensemble de données est lié au marketing direct campagnes d'une institution bancaire portugaise; contient 4521 instances et 17 descripteurs.</li> <li>• <b>Méthodologie:</b> Une phase de prétraitement est d'abord mise en œuvre pour équilibrer la distribution des données, ensuite, forêt aléatoire, SVM, ANN, Naïve Bayes des algorithmes de classification sont appliqués, et leur évaluation est effectuée en utilisant Gmean<sup>17</sup> comme un évaluateur métrique.</li> </ul>
<b>Term Deposit Subscription Prediction Using Spark<sup>18</sup> MLib and ML Packages<sup>3</sup></b>	<ul style="list-style-type: none"> <li>• <b>Données:</b> L'ensemble de données est obtenu à partir du site de machine learning UCI<sup>19</sup>. Cet ensemble de données est lié au marketing direct campagnes d'une institution bancaire portugaise.</li> <li>• <b>Méthodologie:</b> Dans cette études les différentes étapes d'exploration de données ont été simulés en utilisant le langage Python<sup>20</sup> et les différentes librairies qui comporte tel que MLib<sup>21</sup> et bien d'autres.</li> </ul>

<sup>11</sup> Ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence

<sup>12</sup> Algorithme d'apprentissage supervisé de classifieurs binaires.

<sup>13</sup> Le prétraitement des données est une étape importante du processus d'exploration de données, c'est dans phase qu'en nettoyant les données en éliminant les intrus

<sup>14</sup> En utilisant des algorithmes et des outils informatiques

<sup>15</sup> En utilisant la matrice de confusion et la validation croisée

<sup>16</sup> Université de Californie à Irvine

<sup>17</sup> Moyenne géométrique : qui indique la tendance centrale ou à la valeur typique d'un ensemble de nombres en utilisant le produit de leurs valeurs (par opposition à la moyenne arithmétique qui utilise leur somme). La moyenne géométrique est définie comme étant la nème racine du produit de n nombres.

<sup>18</sup> Il s'agit d'un ensemble d'outils et de composants logiciels structurés selon une architecture définie. Développé à l'université de Californie

<sup>19</sup> Université de Californie à Irvine

<sup>20</sup> Un langage de programmation interprété, Il favorise la programmation impérative structurée, fonctionnelle et orientée objet.

<sup>21</sup> Une librairie de Python avec les principaux algorithmes d'apprentissage et méthodes statistique.

<p><b>Predicting customer response to bank direct telemarketing campaign<sup>4</sup></b></p>	<ul style="list-style-type: none"> <li>• <b>Données:</b> L'ensemble de données est obtenu à partir du site de Machine Learning UCI<sup>22</sup>. Cet ensemble de données est lié au marketing direct campagnes d'une institution bancaire portugaise; contient 4521 instances et 17 descripteurs.</li> <li>• <b>Méthodologie:</b> L'étude applique quatre algorithmes d'apprentissage automatique qui comprend le réseau neuronal multicouche Perceptron (MLPNN), Arbre de décision (C4.5), Régression logistique et enfin Random Forest (Forêt aléatoire). Pour l'évaluation des modèles a été réalisée à l'aide des précisions de prédiction de ces classificateurs et de la courbe récepteur caractéristique des opérations (ROC<sup>23</sup>).</li> </ul>
<p><b>A Data Mining Approach for Bank Telemarketing Using the Rminer Package and R Tool<sup>5</sup></b></p>	<ul style="list-style-type: none"> <li>• <b>Données:</b> données collectées incluses 17 campagnes exécutées entre mai 2008 et novembre 2010, correspondant à un total de 79354 contacts et 58 attributs pour caractériser chacun d'eux. Pendant ces appels campagnes, une application de dépôt à terme attrayante a été offerte.</li> <li>• <b>Méthodologie:</b> Dans cette étude les différentes étapes de l'exploration de données ont été simulées en utilisant sous R<sup>24</sup>, pour les techniques du Data Mining utilisées: SVM, Arbre de décision et Naïve Bayes, pour l'évaluation des modèles; la fonction Rminergraph qui était utilisé pour tracer les courbes ROC et Lift<sup>25</sup>.</li> </ul>
<p><b>An intelligent system for customer targeting a data mining approach<sup>6</sup></b></p>	<ul style="list-style-type: none"> <li>• <b>Méthodologie:</b> Cette étude est organisée comme suit. Dans la première étape on trouve une explication en détail de GA<sup>26</sup> Dans la deuxième étape on trouve une description de la structure de l'AG / Modèle ANN<sup>27</sup>. Dans la dernière étape, nous trouverons les résultats du modèle GA / ANN.</li> </ul>

### 3.2 LES DIFFERENTES TECHNIQUES UTILISEES

**Tableau 3. Les techniques du Data Mining utilisées dans les articles revus**

Titre de l'article	Les techniques du Data Mining utilisées
<p>The Implementation of Subspace Outlier Detection in K-Nearest Neighbors to Improve Accuracy in Bank Marketing Data<sup>1</sup></p>	<ul style="list-style-type: none"> <li>• <b>KNN<sup>28</sup>:</b> est un algorithme supervisé appliqué pour classer un ensemble des objets ou données basés sur les voisins les plus proches dont la classe est connue. Dans le KNN, il y a trois éléments importants à utiliser pour la classification, y compris les valeurs K, les objets étiquetés, il y a aussi les distances entre les objets. La valeur K dans le KNN est utilisée pour déterminer le numéro du plus proche voisin.</li> </ul>
<p>Visualization and Analysis in Bank Direct Marketing Prediction<sup>2</sup></p>	<ul style="list-style-type: none"> <li>• <b>Random Forest:</b> La forêt aléatoire est un algorithme d'apprentissage supervisé<sup>29</sup> utilisé pour les tâches de classification et de régression. Il se distingue à partir d'arbres de décision par le processus aléatoire de trouver la racine nœuds pour diviser les entités. La forêt aléatoire est efficace dans sa manipulation valeurs manquantes. Sauf si un nombre suffisant d'arbres est généré pour améliorer la précision des prévisions, le problème de sur apprentissage<sup>30</sup> est inconvenient possible de cet algorithme.</li> <li>• <b>SVM:</b> C'est un algorithme d'apprentissage utilisé dans les tâches de régression<sup>31</sup>. Cependant, SVM est préférable dans les tâches de Classification. Cet algorithme est basé sur l'idée suivante: si un classificateur est efficace pour séparer les données convergentes non linéairement séparables points, alors il devrait bien fonctionner sur ceux dispersés. SVM trouve la meilleure ligne de séparation qui maximise la distance entre les hyperplans des frontières de décision.</li> </ul>

<sup>22</sup> Université de Californie à Irvine

<sup>23</sup> ROC : Est une mesure de la performance d'un classificateur binaire, c'est-à-dire d'un système qui a pour objectif de catégoriser des éléments en deux groupes distincts sur la base d'une ou plusieurs des caractéristiques de chacun de ces éléments

<sup>24</sup> R est un langage de programmation et un logiciel libre (Gratuit) destiné aux statistiques et à la science des données soutenu par la R Fondation.

<sup>25</sup> Lift est une mesure de la performance d'un modèle prédictif ou descriptif, mesuré par rapport au modèle du choix aléatoire.

<sup>26</sup> Genetic Algorithm Les algorithmes génétiques appartiennent à la famille des algorithmes évolutionnistes. Leur but est d'obtenir une solution approchée à un problème d'optimisation, lorsqu'il n'existe pas de méthode exacte pour le résoudre

<sup>27</sup> Un réseau de neurones artificiels, est un algorithme dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques.

<sup>28</sup> k nearest neighbors (k plus proches voisins)

<sup>29</sup> L'algorithme apprend ainsi à partir de milliers ou de millions d'exemples étiquetés c-à-d on lui fait entrer des objets étiquetés avec leur label.

<sup>30</sup> Intervient lorsque l'algorithme sur-apprend, autrement dit, lorsqu'il apprend à partir des données mais aussi à partir de patterns (schémas, structures) qui ne sont pas liés au problème, comme du bruit. (Source : <https://dataanalyticspost.com/>)

<sup>31</sup> L'objectif est d'apprendre (autrement dit de trouver) la relation qui lie une variable d'intérêt, de type quantitatif, aux autres variables observées, éventuellement dans un but de prédiction. (Source : <https://dataanalyticspost.com/>)

	<ul style="list-style-type: none"> <li>• <b>Artificial Neural Networks (ANN):</b> Un réseau de neurones artificiels, ou réseau neuronal artificiel, est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques.</li> <li>• <b>Naïve Bayes:</b> Est un classificateur direct et puissant qui utilise le théorème de Bayes. Il prédit la probabilité qu'une donnée appartient à une classe particulière. La classe avec la probabilité la plus élevée est considérée comme la plus probable classe. Cet algorithme suppose que toutes les fonctionnalités sont indépendantes et sans rapport. Le modèle Naive Bayes est simple et facile à construire et particulièrement utile pour les grands ensembles de données. Ce modèle est connu pour surpasser même une classification très sophistiquée méthodes.</li> <li>• <b>KNN<sup>32</sup>:</b> est un algorithme supervisé appliqué pour classer un ensemble d'objets ou données basés sur les voisins les plus proches dont la classe est connue. Dans le KNN, il y a trois éléments importants a utilisé pour la classification, y compris les valeurs K, les objets étiquetés, il y a aussi les distances entre les objets. La valeur K dans le KNN est utilisée pou déterminer le numéro du plus proche voisin.</li> </ul>
<p>Term Deposit Subscription Prediction Using Spark MLlib and ML Packages<sup>3</sup></p>	<ul style="list-style-type: none"> <li>• <b>DecisionTree:</b> Un arbre de décision est un modèle prédictif qui se caractérise par l'auto-formation semi-supervisée. Le modèle est généré à l'aide d'une approche descendante, où l'ensemble de données source est divisé en sous-ensembles à l'aide d'une mesure statistique, souvent sous la forme de l'indice de Gini<sup>33</sup> ou le gain d'informations via l'entropie de Shannon<sup>34</sup>. Cette le processus est appliqué récursivement jusqu'à ce qu'un sous-ensemble ne contienne que des échantillons avec la même classe cible, ou est interrompu par un arrêt prédéfini Critères.</li> <li>• <b>RandomForest:</b> La forêt aléatoire est un algorithme d'apprentissage supervise<sup>35</sup> utilisé pour les tâches de classification et de régression. Il se distingue à partir d'arbres de décision par le processus aléatoire de trouver la racine nœuds pour diviser les entités. La forêt aléatoire est efficace dans sa manipulation valeurs manquantes. Sauf si un nombre suffisant d'arbres est généré pour améliorer la précision des prévisions, le problème de surapprentissage<sup>36</sup> est inconvenient possible de cet algorithme.</li> <li>• <b>Gradient-BoostedTree:</b> L'amplification du gradient est une technique d'apprentissage automatique pour les problèmes de régression et de classification, qui produit un modèle de prédiction sous la forme d'un ensemble de modèles de prédiction faibles, généralement des arbres de décision.</li> </ul>
<p>Predicting customer response to bank direct telemarketing campaign<sup>4</sup></p>	<ul style="list-style-type: none"> <li>• <b>Multilayer Perceptron Neural Network (MLPNN):</b> une technique d'exploration de données robuste à ses fondements sous-jacents ancrés dans l'intelligence artificielle (IA). Il est composé d'une collection dépendante de neurones artificiels qui applique une méthode de calcul connectée. Il est célèbre pour sa prouesse d'approximation fiable dans la prédiction et la classification Tâches. MLPNN complète son excellente approximation capacités avec une grande facilité d'apprentissage, moins de mémoire requise, l'analyse complexe des motifs ainsi qu'une opération rapide dans le temps.</li> <li>• <b>DecisionTree (C4.5):</b> L'algorithme C4.5 est un algorithme de classification supervisé, publié par Ross Quinlan<sup>37</sup>.</li> <li>• <b>LogisticRegression:</b> La régression logistique est un modèle de régression binomiale. Comme pour tous les modèles de régression binomiale, il s'agit de modéliser au mieux un modèle mathématique simple à des observations réelles nombreuses</li> <li>• <b>Random Forest:</b> La forêt aléatoire est un algorithme d'apprentissage supervise<sup>38</sup> utilisé pour les tâches de classification et de régression. Il se distingue à partir d'arbres de décision par le processus aléatoire de trouver la racine nœuds pour diviser les entités. La forêt aléatoire est efficace dans sa manipulation valeurs manquantes. Sauf si un nombre suffisant d'arbres est généré pour améliorer la précision des prévisions, le problème de surapprentissage<sup>39</sup> est inconvenient possible de cet algorithme.</li> </ul>
<p>A Data Mining Approach for Bank Telemarketing Using</p>	<ul style="list-style-type: none"> <li>• <b>Naïve Bayes (NB):</b> Est un classificateur direct et puissant qui utilise le théorème de Bayes. Il prédit la probabilité qu'une donnée appartient à une classe particulière. La classe avec la probabilité la plus élevée</li> </ul>

<sup>32</sup> k nearest neighbors (k plus porches voisins)

<sup>33</sup> Est une mesure statistique permettant de rendre compte de la répartition d'une variable au sein d'une population.

<sup>34</sup> L'entropie de Shannon, due à Claude Shannon, est une fonction mathématique qui, intuitivement, correspond à la quantité d'information contenue ou délivrée par une source d'information.

<sup>35</sup> L'algorithme apprend ainsi à partir de milliers ou de millions d'exemples étiquetés c-à-d on lui fait entrer des objets étiquetés avec leur label.

<sup>36</sup> Intervient lorsque l'algorithme sur-apprend, autrement dit, lorsqu'il apprend à partir des données mais aussi à partir de patterns (schémas, structures) qui ne sont pas liés au problème, comme du bruit. (Source : <https://dataanalyticspost.com/>)

<sup>37</sup> C'est un chercheur en informatique qui fut l'un des premiers à étudier l'exploration de données et la théorie de la décision.

<sup>38</sup> L'algorithme apprend ainsi à partir de milliers ou de millions d'exemples étiquetés c-à-d on lui fait entrer des objets étiquetés avec leur label.

<sup>39</sup> Intervient lorsque l'algorithme sur-apprend, autrement dit, lorsqu'il apprend à partir des données mais aussi à partir de patterns (schémas, structures) qui ne sont pas liés au problème, comme du bruit. (Source : <https://dataanalyticspost.com/>)

the rminer Package and R Tool <sup>5</sup>	<p>est considérée comme la plus probable classe. Cet algorithme suppose que toutes les fonctionnalités sont indépendantes et sans rapport. Le modèle Naïve Bayes est simple et facile à construire et particulièrement utile pour les grands ensembles de données. Ce modèle est connu pour surpasser même une classification très sophistiquée méthodes.</p> <ul style="list-style-type: none"> <li>• <b>DecisionTrees (DT):</b> Un arbre de décision est un modèle prédictif qui se caractérise par l'auto-formation semi-supervisée. Le modèle est généré à l'aide d'une approche descendante, où l'ensemble de données source est divisé en sous-ensembles à l'aide d'une mesure statistique, souvent sous la forme de l'indice de Gini<sup>40</sup> ou le gain d'informations via l'entropie de Shannon<sup>41</sup>. Cette le processus est appliqué récursivement jusqu'à ce qu'un sous-ensemble ne contienne que des échantillons avec la même classe cible, ou est interrompu par un arrêt prédéfini Critères.</li> <li>• <b>Support Vector Machines (SVM):</b> C'est un algorithme d'apprentissage utilisé dans les tâches de régression<sup>42</sup>. Cependant, SVM est préférable dans les tâches de Classification. Cet algorithme est basé sur l'idée suivante: si un classificateur est efficace pour séparer les données convergentes non linéairement séparables points, alors il devrait bien fonctionner sur ceux dispersés. SVM trouve la meilleure ligne de séparation qui maximise la distance entre les hyperplans des frontières de décision.</li> </ul>
An intelligent system for customer targeting a data mining approach <sup>6</sup>	<ul style="list-style-type: none"> <li>• <b>L'algorithme génétique (GA):</b> en tant que technique de recherche globale est utile pour les problèmes d'optimisation complexes où le nombre de paramètres est important et difficile à obtenir.</li> <li>• <b>Artificial Neural Networks (ANN):</b> Un réseau de neurones artificiels, ou réseau neuronal artificiel, est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques.</li> </ul>

#### 4 DISCUSSION

Tout au long de cette revue de littérature on a pu extraire en sus des nouvelles connaissances des nouvelles approches et méthodes pour réussir sa stratégie numérique. Dans l'ensemble des articles qu'on choisit la base de données qui était utiliser appartient à une banque portugaise ce qui a rendu ce travail assez robuste en termes d'unicité des données et de multiplicités de méthodes utilisées. Par ailleurs les résultats obtenus on révéler que certains algorithmes ou techniques du Data Mining peuvent servir pour réussir et monter une stratégie du marketing assez pertinente.

En effet le KNN en alliance avec SOD a donné des résultats satisfaisants avec une précision de 94%, en outre le Naïve Bayes a été préféré selon l'étude qu'a été mener dans le deuxième article et ce pour sa simplicité et sa maintenabilité, alors pour la quatrième étude le Random Forest a démontré sa meilleure capacité de prédiction parmi les quatre classificateurs utilisés, dans cette même étude les auteurs ont réalisé une analyse en grappes k-means<sup>43</sup>. À cet égard, les résultats de cette analyse ont révéler que les clients les plus intéressés ceux qui occupaient des postes de direction et qui étaient mariés et avec un niveau minimum d'éducation étant secondaire. Également l'analyse dans cette quatrième étude a montré que le mois de septembre s'est avéré être le mois « le plus fertile » pour les clients qui souscrivent au dépôt à terme. En ce qui nous concerne les logiciels utilisés les résultats du cinquième montrent qu'il est possible d'analyser les données bancaires de télémarketing avec des outils open source (Gratuit), notamment en utilisant l'environnement R et paquet Rmine

Dans l'ensemble on peut conclure que les techniques du Data Mining peuvent renforcer et accélérer le processus du marketing en adoptant les méthodes adéquates et en explorant les données nécessaires pour réussir sa stratégie numérique.

#### 5 CONCLUSIONS

Le marketing digital vise l'atteinte des objectifs du marketing traditionnel et ce en utilisant le géant réseau d'internet et les dernières technologies numériques. Il ne remet pas en cause les principes fondamentaux du marketing tels que segmentation, ciblage et positionnement, mais prend en compte le développement massif des nouveaux usages numériques par les individus et celui des nouvelles technologies. Internet et réseaux sociaux sont devenus les tunnels les plus populaires partout dans le

<sup>40</sup> Est une mesure statistique permettant de rendre compte de la répartition d'une variable au sein d'une population.

<sup>41</sup> L'entropie de Shannon, due à Claude Shannon, est une fonction mathématique qui, intuitivement, correspond à la quantité d'information contenue ou délivrée par une source d'information.

<sup>42</sup> L'objectif est d'apprendre (autrement dit de trouver) la relation qui lie une variable d'intérêt, de type quantitatif, aux autres variables observées, éventuellement dans un but de prédiction. (Source : <https://dataanalyticspost.com/>)

<sup>43</sup> Le partitionnement en k-moyennes est une méthode de partitionnement de données et un problème d'optimisation combinatoire. Étant donné des points et un entier k, le problème est de diviser les points en k groupes, souvent appelés clusters, de façon à minimiser une certaine fonction

monde, et toutes les organisations se doivent d'y construire une présence cohérente sur les différents types de médias afin de recruter des nouveaux clients en adoptant les meilleures stratégies numériques.

## REFERENCES

- [1] Anggoro, d. A. The implementation of subspace outlier detection in k-nearest neighbors to improve accuracy in bank marketing data. *Int. J. Emerg. Trends eng. Res.*8, 545–550 (2020).
- [2] Abu-srhan, a., alhammad, b., al, s. & al-sayyed, r. Visualization and analysis in bank direct marketing prediction. *Int. J. Adv. Comput. Sci. Appl.*10, (2019).
- [3] Hung, p. D., hanh, t. D. & tung, t. D. Term deposit subscription prediction using spark mllib and ml packages. In proceedings of the 2019 5th international conference on e-business and applications - iceba 2019 88–93 (acm press, 2019).  
Doi: 10.1145/3317614.3317618.
- [4] Asare-frempong, j. & jayabalan, m. Predicting customer response to bank direct telemarketing campaign. In 2017 international conference on engineering technology and technopreneurship (ice2t) 1–4 (ieee, 2017).  
Doi: 10.1109/ice2t.2017.8215961.
- [5] Drucker peter. *Management, task, responsibilities and practices.* (edition harper & row, 1973).
- [6] Eric vernette. *L'essentiel du marketing : marketing fondamental.* (édition d'organisation, 1998).
- [7] Emer mestiri. *Les techniques d'exploitation de données (data mining);*  
[https://www.sas.com/content/dam/sas/en\\_ca/user%20group%20presentations/montreal-user-group/mestiri-mining-fall2015.pdf](https://www.sas.com/content/dam/sas/en_ca/user%20group%20presentations/montreal-user-group/mestiri-mining-fall2015.pdf).
- [8] Jaques lendrevie, denis indon. *Théories et pratiques de marketing.* (edition dolez, 1997).
- [9] Kim, y. & street, w. N. An intelligent system for customer targeting: a data mining approach. *Decis. Support syst.*37, 215–228 (2004).
- [10] Lambin jean-jacque et chantal de moerloose. *Le marketing stratégique et opérationnel.* (dunod, 2008).
- [11] Moro, s., cortez, p. & laureano, r. M. S. A data mining approach for bank telemarketing using the rminer package and r tool. 24.
- [12] Philip kotler, bernard dubois. *Marketing et management.* Édition: pearson, 2005,.
- [13] Philip kotler et gary armstrong. *Principes de marketing.* Édition: pearson, 2013.
- [14] Pierre louis dubois, alain jolibert. *Le marketing : fondements et pratiques.* (2ème édition, economica, 1992).
- [15] Pierre-alain four. *Des données au savoir : big data et data mining;*  
<https://www.millenaire3.com/content/download/3160/53603>.
- [16] Scheid, françois, fontugne, willy & vaillant, renaud. *Le marketing digital 'développer sa stratégie numérique'.* (2019).