

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA

Faculté des Mathématiques et
de l'Informatique
Département d'Informatique

N° :



DOMAINE : Mathématiques
et Informatique

FILIERE : Informatique

OPTION : IDO

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par: Benkaihoul Ikram

Hamlaoui Menel

Intitulé

Un modèle d'analyse du Comportement Touristique

Etude de cas : Wilaya de M'sila

Soutenu devant le jury composé de :

Mokhtari Rabah

Université de M'sila

Président

Tahar Mehenni

Université de M'sila

Rapporteur

Chalabi Baya

Université de M'sila

Examineur

Année universitaire : 2022 / 2023

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA

Faculté des Mathématiques et

de l'Informatique

Département d'Informatique

N° :



جامعة محمد بوضياف - المسيلة
Université Mohamed Boudiaf - M'sila

DOMAINE : Mathématiques

et Informatique

FILIERE : Informatique

OPTION : IDO

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par: Benkaihou Ikram

Hamlaoui Menel

Intitulé

Un modèle d'analyse du Comportement Touristique

Etude de cas : Wilaya de M'sila

Soutenu devant le jury composé de :

Mokhtari Rabah

Université de M'sila

Président

Tahar Mehenni

Université de M'sila

Rapporteur

Chalabi Baya

Université de M'sila

Examineur

Année universitaire : 2022 / 2023

Dédicaces

Nous dédions se mémoire :

*A nos parents pour leurs encouragements et leur
support.*

A tous nos frères et sœurs.

Et à tous ceux qui nous ont aidé de près ou de loin.

*A nos amis qui nous ont encouragé et aidé pour
réaliser cet humble travail.*

Remerciements

**Nous voulions prendre un moment pour exprimer
notre gratitude nos sincères remerciements à tous ceux
qui ont contribué à la réalisation de notre mémoire de
master.**

**Remercier aussi notre encadreur Mr Tahar Mehenni
pour son encadrement.**

**Comme nous tenons à remercier les membres du jury
(Mokhtari Rabah, Chalabi Baya) d'avoir accepté
D'examiner et juger notre travail.**

**Un grand merci à notre famille et à nos amis pour
leur soutien et leur encouragement.**

Table des matières

Table des matières	6
Liste des figures	9
Liste des tableaux	10
Liste des abréviations	11
Introduction Générale.....	12
CHAPITRE 1 Tourisme et comportement touristique.....	14
Introduction	14
1. Définition du tourisme	15
2. Les types du tourisme.....	16
3. Rôle, importance du tourisme	17
3.1. Rôles du tourisme.....	17
3.2. L'importance de tourisme.....	17
4. Critères de classification du tourisme	18
5. Services d'appui au tourisme	18
6. Comportement touristique.....	18
6.1. Définition du touriste	18
6.2. types de touristes:	19
6.3. La destination touristique	20
6.4. Comportement touristique:.....	22
Conclusion	24
CHAPITRE 2 Les modèles prédictifs	26
Introduction	26
1. Définition	26

1.1. L''utilité le data mining aujourd'hui:	27
1.2. Concepts clés du data mining.....	28
2. La modélisation prédictive	29
2.1. Les domaines d'application de La modélisation prédictive.....	29
2.2. Comment se déroule un processus type de modélisation prédictive?	30
2.3. Les modèles d'analyse prédictive	31
3-Quelques meilleurs outils pour faire en data mining	39
3.1-Python:	39
3.2-Le langage R	40
3.3-Tanagra	40
Conclusion	41
CHAPITRE 3 Etat de l'art	42
Introduction	42
1.Travail 01: Next Location Recommendation with Attention Mechanism by Mining the Spatio-Temporal Relationship between Visited Locations.....	43
2.Travail 02:Tourist Behavior Pattern Mining Model Based on Context.....	44
3.Travail 03: Exploiting Social Images to Understand TouristBehaviour.	44
4.Travail 04: A big data analytics method for tourist behaviour analysis.....	45
Conclusion	46
CHAPITRE 4 : Approche proposée, Résultats et Discussion.....	47
Introduction	47
1.Lieu d'étude : (wilaya de M'sila)	47
1.1. Définition	47
1.2. Zones et sites touristiques	48
2-Méthodologie et préparation des données	48
3-Les outils et détails d'implémentation	50
3.1-Langage de programmation	50
3.3 Matrice de Confusion.....	52
4-L'apprentissage du modèle proposé (Classification).....	53
4.1-Présentation des ensembles d'apprentissage utilisés	53

Table des matières

4.2- L'importation et le partitionnement de la base de données	54
4.3. Résultats d'évaluation des approches d'apprentissage automatique sur les donnée.....	57
Conclusion	58
Conclusion générale.....	59
Bibliographie.....	60
Résumé.....	62
Abstract	63
:ملخص.....	64

Liste des figures

Figure 2.1: Logo Data Maining.....	27
Figure 2.2: la modélisation prédictive.....	31
Figure2. 3 : Schéma de model de classification.....	33
Figure 2.4: Construction du modèle.....	33
Figure 2.5: utilisation du modèle.....	34
Figure 2.6: l'arbre de décision.....	35
Figure 2.7: Exemple de l'arbre de décision.....	35
Figure 2.8: Méthode à noyau SVM.....	37
Figure 2.9: Modèle de Régression.....	37
Figure 2.10: Une relation linéaire simple entre deux variables.....	39
Figure 2.11: logo python.....	39
Figure 2.12: logo R.....	40
Figure 2.13:logo Tanagra.....	40
Figure 4.1:nombre de touristesdans M'sila (2012-2022).....	48
Figure 4.2:Extrait de la Base de données en Excel.....	49
Figure 4.3:Logo de l'IDE Jupyter.....	51
Figure 4.4: Logo de bibliothèque Sklearn.....	52
Figure 4.5: Un groupe d'apprentissage pour prédire le comportement touristique (BDD1).	53
Figure 4.6: Graphique de la variable dépendante.....	54
Figure 4.7: L'importation de l'ensemble d'apprentissage et la définition des variables.	54
Figure 4.8: Code de Partitionnement de l'ensemble d'apprentissage.....	54
Figure 4.9: modèle Arbre de décision et matrice de confusion.....	55
Figure 4.10: Les résultats de matrice de confusion.....	55
Figure 4.11: un code pour dessinée l'Arbre de Décision.....	56
Figure 4.12: Un extrait du graphe d'Arbre de Décision BDD.....	56
Figure 4.13 : modèle Naïve Bayes et matrice de confusion.....	57
Figure 4.14: Les résultats de matrice de confusion.....	57
Figure 4.15 : modèle SV M et matrice de confusion.....	58
Figure 4.16: Les résultats de matrice de confusion.....	58

Liste des tableaux

Tableau 3.1: Etat de l'art.....	46
Tableau 4.1: Exemple de Matrice de Confusion.....	52
Tableau 4.2: Les résultats obtenus en Arbre de Décision.....	55
Tableau 4.3: Résultats obtenus par le modèle Naïve bayes.....	57
Tableau 4.4: Résultats obtenus par SV M.....	58
Tableau 4.5: Pourcentage de Accourcie pour BDD	59

Liste des abréviations

IA : Intelligence Artificiel

DM : Data Mailing

DT: Décision Tree

SVM : Support Vector Machines

SVC : Support Vector Classification

LDA : Analyse discriminante linéaire

ACC : Accuracy

RBF : Radial Basis Function

BDD : Base de données

Introduction Générale

Le tourisme est une industrie mondiale florissante qui comprend le déplacement de personnes d'un endroit à un autre à diverses fins, comme les loisirs, les affaires, l'éducation ou l'exploration culturelle. Il joue un rôle important dans le développement économique, social et culturel des pays du monde entier. Le tourisme permet aux gens de découvrir différentes destinations, de s'immerger dans diverses cultures et de créer des expériences mémorables.

Le tourisme est une activité qui englobe un large éventail d'expériences, de comportements et d'interactions entre les individus et leur environnement. Il implique le déplacement de personnes d'un lieu à un autre, que ce soit à des fins de loisirs, de découverte culturelle, de détente ou d'affaires. Le comportement touristique se réfère aux actions, attitudes et motivations des touristes pendant leur voyage.

L'analyse du comportement touristique est une discipline qui vise à comprendre les attitudes, les motivations et les actions des touristes lorsqu'ils planifient, choisissent et participent à des activités touristiques. Elle permet de recueillir des informations précieuses sur les préférences, les besoins et les comportements des touristes, ce qui est essentiel pour les professionnels du tourisme, les agences de voyage, les destinations touristiques et d'autres acteurs de l'industrie.

Un modèle d'analyse du comportement touristique est un outil conceptuel qui aide à structurer et à analyser les différentes dimensions du comportement des touristes. Il peut être utilisé pour identifier les facteurs clés qui influencent les décisions des touristes, prédire leurs comportements futurs, segmenter les marchés touristiques et concevoir des stratégies de marketing adaptées.

Ce modèle comprend généralement plusieurs éléments. Tout d'abord, il prend en compte les facteurs socio-démographiques tels que l'âge, le sexe, le niveau d'éducation et le revenu, qui influencent les préférences et les motivations des touristes. Ensuite, il examine les variables psychologiques telles que les besoins, les motivations, les attitudes et les perceptions des touristes, qui jouent un rôle important dans leurs décisions de voyage.

L'utilisation d'un modèle d'analyse du comportement touristique permet aux professionnels du tourisme de mieux comprendre les besoins et les attentes des touristes, d'adapter leurs produits et services en conséquence et de développer des stratégies de marketing plus efficaces. Cela peut contribuer à améliorer l'expérience touristique globale, à stimuler la satisfaction des touristes et à favoriser le développement durable de l'industrie du tourisme

Nous proposerons, dans ce mémoire une approche basée sur l'apprentissage automatique supervisé, pour prédire le comportement touristique. Ce modèle a été évalué sur la base de certaines des zones touristiques de M'sila qui ont été visitées par les touristes au cours des dernières années.

Ce mémoire est organisé comme suit :

Chapitre 1 : Tourisme et Comportement touristique.

Chapitre 2 : Les modèles prédictifs.

Chapitre 3 : Etat de l'art.

Chapitre 4 : Approche Proposée, Résultats et Discussion

Enfin, le mémoire est terminé par une conclusion générale.

CHAPITRE 1 Tourisme et comportement touristique

Introduction :

Le tourisme est l'une des nécessités de la civilisation moderne en raison de ses effets positifs et un rôle distinct dans le soutien L'économie nationale, la réduction du chômage et la revitalisation du mouvement commercial entre les pays, parce que nous ne pouvons pas imaginer Avoir un pays civilisé sans hôtels et tourisme et fournir divers biens et services touristiques qui peuvent Les besoins, les désirs et les goûts des touristes ont lieu grâce à la présence d'installations touristiques qui reflètent le modèle touristique.

Il peut satisfaire les tendances et les désirs des touristes comme ils le souhaitent et aspirent à Sur la base de cela, nous trouvons des pays intéressés par une ligne touristique particulière en fonction de l'environnement touristique distinctif C'est parce qu'ils ont des attractions, qu'il s'agisse de témoignages naturels, civilisationnels ou historiques, et ainsi de suite.

Nous pouvons dire que l'industrie touristique a été et continue d'être en mouvement avec le reste de la Autres industries.

D'où la nécessité de comprendre de nombreux facteurs importants qui sont directement liés au mouvement Le tourisme est la principale cible et le principal moyen touristique, ce qui nécessite une sensibilisation L'importance de l'information intégrée sur les touristes en termes de besoins, désirs et goûts reflétés dans Ses comportements.

L'étude des comportements touristiques est très sensible car elle est liée à l'avenir de l'industrie touristique pour la conservation Sur la marque de l'installation touristique et assurer le flux continu de touristes vers la zone touristique Le modèle touristique ainsi que d'autres facteurs influençant le comportement du touriste, c'est ce qui rend le touriste Regardez le modèle touristique avec plus d'un regard et idée pour obtenir ses désirs devant les multiples variables dans le modèle touristique, qui oriente souvent le comportement du touriste et l'attire à lui.

1. Définition du tourisme :

Le tourisme est l'acte de voyager pour le plaisir ou les affaires vers une destination en dehors de son environnement habituel et d'y séjourner pour une durée limitée. C'est une industrie multifacette qui englobe un large éventail d'activités, notamment les transports, l'hébergement, la restauration, les attractions, les événements et les services.

Le tourisme est un phénomène qui surgi depuis que Dieu a créé la terre, et maintenant est devenu un phénomène social, culturel qui reflète sur le développement de la civilisation des peuples du monde, et c'était quelques personnalités du monde, qui ont conduit à une augmentation du tourisme, de leur rôle et de leur contribution dans la croissance des voyages ainsi que les développements technologiques de différentes manières.

Au fil des décennies, le tourisme a connu un essor continu et s'est diversifié de plus en plus, au point de devenir un des secteurs économiques à la croissance la plus rapide du monde. Le tourisme moderne est étroitement lié au développement, et il englobe un nombre grandissant de nouvelles destinations. Cette dynamique en fait un moteur essentiel du progrès socioéconomique.

Aujourd'hui, le volume d'affaires du secteur touristique égale, voire dépasse celui des industries pétrolière, agroalimentaire ou automobile. Le tourisme est désormais un des grands acteurs du commerce international et, en même temps, il constitue une des principales sources de revenus de beaucoup de pays en développement. Cette croissance va de pair avec l'accentuation de la diversification et de la concurrence entre les destinations.

Le tourisme c'est un ensemble d'activités relié avec les voyages et les déplacements hors ville ou hors pays pour différent besoins.

On définit le tourisme comme une activité humaine, qui est basée sur un déplacement, Un changement de place et par extension géographique un changement d'habitat. Cette caractéristique qui le différencie des loisirs pour lesquels le déplacement n'est pas vraiment une obligation et une condition nécessaire à leur réalisation. Faire du tourisme signifie quitter temporairement son lieu de vie habituel pour aller ailleurs, dans un ou des lieux situés hors de son environnement quotidien.

C'est le fait de quitter son domicile, pour des raisons personnelles, pour une durée supérieure à 24 heures. Ce qui peut impliquer la consommation d'une nuitée auprès d'un hôtelier et éventuellement la réservation de titre de transport. Initialement et uniquement rattaché aux loisirs et à la santé, le tourisme englobe désormais également l'ensemble des

activités économiques auxquelles la personne fait appel lors d'un déplacement inhabituel (transports, hôtels, restaurants...).

2. Les types du tourisme :

Afin de répondre à la diversité de la demande et aux goûts de chacun, le tourisme a été segmenté en différents secteurs. Les principaux sont les suivants :

Le tourisme bleu (ou littoral) :

On appelle tourisme bleu le tourisme ayant pour destination les vacances au bord de la mer. C'est la première destination touristique.

Le tourisme vert (ou rural) :

Le tourisme vert correspond aux vacances à la campagne, l'espace rural. C'est la deuxième destination pour les vacancesété.

En matière d'hébergement, les gîtes et les chambres d'hôtes connaissent le plus vif succès. Viennent ensuite les hôtels regroupés sous le label.

Les activités les plus pratiquées en espace rural sont la pêche, le vélo et la randonnée, suivies d'autres loisirs de nature comme le golf, l'équitation, l'escalade.

Le tourisme montagnard (ou blanc en hiver) :

C'est le tourisme à la montagne.

Le tourisme urbain (ou culturel) :

Ce sont les voyages touristiques effectués vers la ville. Le nombre de vacances en ville a augmenté de 50 % soit une croissance deux fois plus forte que le total des vacances.

Les pratiques culturelles sont les activités principales des touristes en ville

Le tourisme de santé :

Tourisme de santé non médical c'est La fréquentation des stations thermales et Les centres de thalassothérapie.

Le tourisme religieux :

Les pèlerinages sont une des plus vieilles formes de migration touristique

Le tourisme d'affaires :

Le tourisme d'affaires regroupe 4 activités: les congrès et conventions d'entreprises, les foires et les salons, les voyages de motivation ou « incentives » et séminaires d'entreprises et les voyages d'affaires individuels.

Le tourisme durable:

Le tourisme durable représente la forme de développement, d'aménagement, d'activité touristique qui respecte l'environnement, préserve à long terme les ressources naturelles

et culturelles. Il est socialement et économiquement durable et équitable.

Le tourisme de découverte économique:

Le tourisme de découverte économique peut être défini comme la découverte par le public d'un site présentant un savoir-faire appartenant au passé, au présent ou à l'avenir.

Trois pôles peuvent être mis en évidence : le tourisme d'entreprises en activité, le tourisme de patrimoine industriel et le tourisme scientifique.

3. Rôle, importance du tourisme :

3.1. Rôles du tourisme:

Dans tous pays, le secteur du voyage et du tourisme contribue de manière importante à l'économie nationale, et c'est un outil puissant de développement territorial car il permet de créer du travail décent et de la richesse dans tout le pays.

3.2. L'importance de tourisme :

Le tourisme plus qu'un phénomène est devenu une industrie qu'un pays sur la planète, ce dernier fait aujourd'hui de nombreuse recherche dans le champ de la science sociale. Quelque soit la discipline (histoire, sociologie, anthropologie ou géographie), le tourisme n'est plus seulement appréhendé en terme de flux, de couts, apport,...mais bien comme un système complexe qui doit être observé sous des angles différents et complémentaires. Le tourisme joue un rôle très important dans les différents secteurs :

Economique :

- Permettre la création d'emplois (01lit d'hôtel/0.5 emploi)
- Développement régional
- Favorise un aménagement de territoire plus équilibré
- Augmente les ressources monétaires

Social :

- Permet d'échange culturel entre les individus
- Permet d'évader d'un environnement stressant exigeant et pollué

Politique :

- Création d'un mouvement d'affaire intense entre les pays
- Favorise la connaissance des pays aux étrangers et leur donne une importance au niveau international.

Culturel :

- Découverte de nouveaux horizons culturel et historique et tradition des pays et des peuples.

- La mise en valeur des potentialités du pays en matière de patrimoine historique et architectural.

4. Critères de classification du tourisme:

Le lieu : Tourisme balnéaire, tourisme urbain, tourisme climatique (montagne, désert ...etc.).

La durée : Tourisme organisé, Tourisme « short breaks », tourisme court durée.

La taille de groupe : Tourisme de masse, tourisme ambiant.

Le budget : Tourisme d'élite (luxe), tourisme social.

L'âge : Tourisme juvénile (des jeunes), tourisme fitness (3eme âge).

La saison : Tourisme hivernal, Tourisme estival.

Le mode d'hébergement : Hôtellerie, Village de vacance, Résidence secondaire, Camping.

5. Services d'appui au tourisme:

Le style touristique se distingue par l'ensemble des services qui le soutiennent et représentent l'identité de l'activité touristique et de ses activités. Étant donné que ces services sont fournis en réponse à un segment spécifique des touristes ciblés qui présentent des caractéristiques Similaire dans le comportement et le comportement, le but de pratiquer des activités touristiques de toutes sortes, nous trouvons Il existe une différence dans le niveau et la qualité des services fournis dans chaque établissement touristique, en fonction du degré de classification Le tourisme et le but pour lequel l'installation a été conçue. Autrement dit, la qualité des services fournis est liée à la qualité de l'activité touristique

Ce qui peut être fourni et ce que les touristes souhaitent avoir, et il existe plusieurs principaux types de services de soutien qui sont principalement fournis dans les services touristiques

6. Comportement touristique:

6.1. Définition du touriste:

Comme il n'y a aucun accord complet sur le sens du terme tourisme il n'y a pas non plus d'accord complet sur la définition du touriste, d'après les nations unies, « un voyageur est une

personne qui se déplace entre deux ou plusieurs pays ou entre localités dans son pays de résidence habituel.

Depuis 1963, les statistiques de L'OMT comprennent les définitions suivantes :

Le visiteur : toute personne qui se rend dans un pays autre que celui où elle réside, pour toute raison que celle d'y exercer une profession rémunérée, il existe deux catégories de visiteur

Les touristes : visiteur temporaire, séjournant au moins 24 heures dans le pays visité (donnant lieu à une nuitée dans un moyen d'hébergement du pays) et dont les motifs du voyage peuvent être groupés en :-loisirs-affaires- Familles

Les excursionnistes : (les visiteurs de la journée) visiteur dont le séjour ne dépasse pas 24 heures ou il ne dépasse pas la nuit dans un moyen d'hébergement collectif ou privé
L'organisation mondiale du tourisme (OMT) essaye de faire appliquer cette définition par l'ensemble des pays concernés afin d'homogénéiser les sources statistiques.

L'organisation mondiale du tourisme définit le touriste comme suit : « toute personne qui se rend dans un pays (tourisme international) ou dans un lieu situé dans son pays de résidence (tourisme interne) mais autre que celui correspondant à son environnement habituel et dont le motif principal de visiter ».

6.2. types de touristes:

Selon le nombre : Les touristes sont répartis comme suit :

Tourisme individuel : dans lequel l'individu communique directement avec les différentes entreprises touristiques.

Tourisme de groupe ou voyages : par l'intermédiaire d'entreprises de tourisme spécialisées.

Selon l'âge :

Touristes pionniers : Ce tourisme est représenté dans les tranches d'âge de 7 à 14 ans, et c'est une étape éducative, au cours de laquelle les enfants acquièrent des connaissances et des compétences comportementales, à travers les écoles et les associations caritatives qui se coordonnent avec les entreprises touristiques pour organiser divers voyages.

Jeunes touristes : Ce type cible la tranche d'âge entre 15-35 ans, et se caractérise par la recherche de vie sociale, d'autonomie, et de mixité avec les autres en faisant des croisières ou des séjours sportifs pour les jeunes.

Touristes matures : ciblant la tranche d'âge entre 35 et 55 ans, et ce type de touriste recherche souvent le repos et la détente ou s'échappe de l'atmosphère de travail routinier et

profite du temps, en particulier à la campagne, sur les plages de la mer, à la montagne et au calme lieux.

Selon la durée du séjour : C'est comme suit :

Touristes saisonniers : il est lié à une saison spécifique, c'est-à-dire que le touriste passe une certaine saison dans un lieu, c'est-à-dire la période de résidence. Passer d'un point à trois points.

Touristes à la journée : Cela prend généralement quelques jours, de deux jours à une semaine, que le touriste passe dans le cadre d'un programme spécifique à l'avance, souvent en fin de semaine ou à l'occasion de fêtes nationales ou de jours fériés.

Touristes de passage : lors du déplacement des touristes par voie terrestre, par bus touristiques, ou par tourisme de passage Se produit lors d'un déplacement en avion, comme un accident d'avion à son aéroport.

Selon le périmètre géographique :

Le touriste local : signifie le mouvement des individus à l'intérieur de l'État lui-même, c'est-à-dire le mouvement psychologique des citoyens de l'État à l'intérieur de celui-ci Ce type de tourisme, de par sa taille, nécessite des services et des prix variés et encourageants pour les citoyens, ce qui est considéré comme l'un des types de tourisme les plus importants.

Touristes étrangers: Ils désignent le mouvement des touristes étrangers dans un pays, et ce type de tourisme est le plus demandé Les pays du monde entier l'encouragent à obtenir des devises fortes. Le tourisme étranger ne peut réussir sans un tourisme intérieur réussi.

6.3. La destination touristique:

6.3.1-Définition le destination:

Une destination, à la différence de l'origine ou du marché, réfère à l'endroit où le touriste a l'intention de passer son temps à l'extérieur de la maison. Cette unité géographique visitée par des touristes peut être un centre indépendant, un village ou une ville, une région ou une île ou un pays. Par ailleurs, une destination peut être un seul emplacement, un ensemble de plusieurs destinations dans le cadre d'un circuit ou même une destination en déplacement telle qu'une croisière par exemple.

- La destination touristique est la place où les individus passent leurs vacances.
- La destination touristique est la place où les individus passent leurs vacances.

- La destination touristique se compose de cinq composantes:
- Les attractions naturelles d'un site;
- Les facilités et les services;
- La facilité d'accès pour ce site;
- L'image construite afin d'attirer les touristes vers ce site;
- Le coût total de vacances.

6.3.2-Choisissez le destination touristique:

Il existe de nombreuses possibilités pour choisir une destination touristique et cela dépend de plusieurs critères ainsi que des sources d'information

1- Les critères de sélection:

En avoir envie

En avoir envie est le premier critère qu'un touriste doit avoir pour choisir une destination car c'est à partir ce critère que se déclenche l'envie de planifier son séjour et voir d'autres critères.

La durée du voyage

Le temps dont on dispose est probablement le critère le plus fondamental. Étant donné le prix élevé des billets d'avion, mieux vaut rentabiliser son investissement et passer plus de temps là où ça coûte le plus cher de se rendre. Il faut aussi considérer qu'un décalage horaire.

Important peut affecter le voyageur pendant plusieurs jours. Le touriste a intérêt à prendre ce facteur en considération. Autrement dit, si celui-ci ne dispose que de deux semaines serait sage de lésiner sur les distances et de se limiter quant au nombre de fuseaux horaires à franchir.

La période de l'année

La période de l'année que le voyageur choisit pour se déplacer est aussi d'une importance considérable. Le climat et la durée du jour changent selon les saisons, et avant de fixer son choix, le voyageur doit s'informer soigneusement.

Le budget

Les moyens financiers dont on dispose influencent considérablement le choix d'une destination. Cependant, la croyance populaire voulant que plus c'est loin, plus c'est cher, n'est pas toujours vraie.

Dans la très grande majorité des pays en voie de développement, le coût de la vie est ridiculement bas pour l'Occidental moyen. Ça coûte évidemment plus cher de s'y rendre que de s'envoler vers un autre pays occidental mais, une fois arrivé, on y vit très à l'aise avec un bien plus petit budget. Un pays occidental se révélera possiblement plus économique pour un court séjour alors que pour un long séjour, une destination plus exotique s'avérera plus avantageuse.

Les capacités physiques

À moins de s'être concocté un projet très «sportif», voyager demande rarement des capacités supérieures à celles nécessaires pour la pratique des activités quotidiennes. Il faut surtout pouvoir marcher, et beaucoup marcher.

Cependant, certains pays aux réseaux routiers déficients peuvent demander une résistance à la fatigue supérieure à la moyenne quand on y voyage au ras du sol.

L'expérience

Il est toujours préférable de graduer les contrastes et les difficultés, d'un voyage à l'autre, et de prendre de l'expérience ailleurs avant d'aborder des pays aux cultures dramatiquement différentes, comme l'Inde, sinon on risque de rater les plus belles et les plus enrichissantes expériences de sa vie et de se dégoûter à jamais du voyage.

2- Des sources d'information:

Les touristes s'appuient souvent sur plusieurs sources d'information lorsqu'ils prennent des décisions de voyage, y compris des sources via :

- l'Internet
- Recommandations d'amis et de la famille
- agents de voyage
- Consulter les forums de voyage

6.4. Comportement touristique:

Le comportement touristique fait référence aux activités, actions et décisions d'individus ou de groupes de personnes qui se livrent à des activités touristiques. Cela implique divers aspects tels que le choix de la destination, la sélection du mode de déplacement, la recherche d'informations, les préférences d'hébergement, la sélection des attractions et des activités et le comportement de consommation. Le comportement des touristes est influencé par divers facteurs, notamment les caractéristiques personnelles, les données sociodémographiques, les motivations, les expériences et les facteurs situationnels.

Les touristes ont différentes motivations pour voyager, notamment la détente, l'aventure, les expériences culturelles, l'évasion, la socialisation et l'éducation. Ces motivations ont un impact sur les types de destinations et d'activités qu'ils choisissent, ainsi que sur leurs habitudes de voyage et leurs habitudes de dépenses. Le comportement des touristes est également influencé par leurs expériences passées, leurs antécédents culturels et leur socialisation.

La recherche d'informations est un aspect critique du comportement des touristes. Les touristes s'appuient souvent sur plusieurs sources d'information lorsqu'ils prennent des décisions de voyage, y compris des sources en ligne, des recommandations d'amis et de famille et des agents de voyage. Ils s'engagent également dans diverses activités avant le voyage, telles que la lecture de guides de voyage, la visite de forums de voyage et la vérification d'avis sur des sites Web de voyage.

Les préférences et la sélection en matière d'hébergement sont des aspects importants du comportement des touristes. Les touristes tiennent compte de plusieurs facteurs lors de la sélection de leur hébergement, notamment l'emplacement, le prix, les installations et la réputation de la marque. Ils ont également des préférences différentes pour les types d'hébergement, tels que les hôtels, les centres de villégiature, les locations de vacances et les auberges.

L'attraction et la sélection d'activités sont des composantes essentielles du comportement touristique. Les touristes choisissent les attractions et les activités en fonction de leurs intérêts, préférences et motivations. Ils tiennent également compte de divers facteurs situationnels, tels que la météo, la disponibilité du temps et l'accessibilité.

Le comportement des touristes a des implications importantes pour la gestion et le marketing des destinations. Comprendre le comportement des touristes aide les destinations à développer des stratégies efficaces pour attirer et retenir les touristes. Il aide également les entreprises à adapter leurs produits et services pour répondre aux besoins et aux préférences des touristes.

6.4.1-Prédire le comportement des touristes et leurs choix de destinations:

L'analyse de données : La collecte et l'analyse de données sur les touristes peuvent fournir des informations sur leur comportement et leurs préférences. Ces données peuvent être obtenues par des enquêtes, des médias sociaux, des requêtes sur les moteurs de recherche et d'autres sources. Les techniques d'apprentissage automatique et de modélisation prédictive

peuvent être appliquées à ces données pour identifier des modèles et faire des prévisions sur le comportement futur.

La collaboration avec les organisations de marketing de destination (DMO) : Les DMO ont souvent des données approfondies sur les tendances du tourisme et peuvent fournir des informations précieuses sur le comportement et les préférences des touristes. La collaboration avec les DMO peut aider les entreprises touristiques à prendre des décisions plus éclairées sur leurs stratégies de marketing et de promotion.

La personnalisation : La personnalisation de l'expérience touristique peut aider les entreprises à mieux comprendre leurs clients et à prédire leur comportement. En collectant des données sur les préférences et les intérêts de chaque individu, les entreprises peuvent offrir des recommandations sur mesure pour les activités, l'hébergement et les destinations.

L'écoute sociale : La surveillance des conversations sur les réseaux sociaux et des avis en ligne peut fournir des informations précieuses sur le comportement et les préférences des touristes. En suivant ce que les gens disent des destinations et des expériences, les entreprises touristiques peuvent identifier des tendances et ajuster leurs stratégies de marketing

Conclusion :

Le tourisme est une industrie mondiale en constante évolution qui englobe de nombreux aspects, y compris les motivations des touristes et leurs comportements lorsqu'ils voyagent. En examinant les tendances actuelles et en prenant en compte les comportements touristiques, il est possible de tirer quelques conclusions importantes.

Tout d'abord, le tourisme continue de croître et de prospérer. Les gens sont de plus en plus attirés par l'idée de découvrir de nouveaux endroits, de vivre de nouvelles expériences et de s'immerger dans différentes cultures. Cela se traduit par une demande accrue de destinations touristiques variées et innovantes.

En ce qui concerne les comportements touristiques, nous constatons une augmentation de l'utilisation des technologies numériques. Les touristes utilisent de plus en plus les applications mobiles, les sites de réservation en ligne et les médias sociaux pour planifier leurs voyages, trouver des informations sur les attractions touristiques et partager leurs expériences avec d'autres.

Le tourisme est en constante évolution et les comportements touristiques reflètent les changements sociaux, environnementaux et technologiques. Les touristes d'aujourd'hui sont plus conscients de leur impact et cherchent des expériences authentiques et durables. Les avancées technologiques ont également un impact important sur la façon dont les voyageurs planifient et vivent leurs voyages. En anticipant ces tendances et en adaptant les offres

touristiques en conséquence, l'industrie du tourisme peut continuer à prospérer tout en répondant aux attentes changeantes des voyageurs .

CHAPITRE 2 Les modèles prédictifs

Introduction :

En apprentissage automatique, un modèle prédictif a pour tâche de prédire, sur la base d'un entraînement sur des données, un résultat (ou une action) pour un problème particulier. Le plus souvent, cette prédiction peut prendre la forme d'une probabilité (ou pourcentage).

Un modèle prédictif de classification sera capable de prédire l'appartenance à une classe pour des données nouvelles sur la base d'un entraînement réalisé auparavant sur un jeu de données étiquetées. Par exemple, un modèle prédictif apprendra à associer le contenu d'un courriel à l'annotation « pourriel » ou « non pourriel » s'il est entraîné sur des centaines, voire des milliers de courriels annotés.

Certains modèles prédictifs exploitent des tendances découvertes à l'intérieur d'ensembles de données historiques (séries chronologiques) afin de prédire un résultat futur.

En prédiction, on distingue les problèmes de régression des problèmes de classement. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression tandis que les problèmes de prédiction d'une variable qualitative sont des problèmes de classification.

Un modèle prédictif peut être un simple classificateur binaire, un modèle de régression pour la prédiction d'un nombre ou quelque chose de plus complexe comme un classificateur multi-classes (multiclass) et multi-étiquettes (multilabel).

La logique prédictive englobe une variété de techniques issues des statistiques, de l'extraction de connaissances à partir de données et de la théorie des jeux. Dans le monde des affaires, des modèles prédictifs exploitent des régularités découvertes à l'intérieur d'ensembles de données historiques et transactionnelles pour identifier les risques ou les occasions d'affaires.

1. Définition :

le data Mining : Extraction d'informations intéressantes (non triviales, implicites, préalablement inconnues et potentiellement utiles) à partir de grandes bases de données. C'est analyser les données pour trouver des patrons cachés en utilisant des moyens

automatiques. C'est un processus non élémentaire de recherche de relations, corrélations, dépendances, associations, modèles, structures, tendances, classes (clusters), segments, lesquelles sont obtenues de grande quantité de données (généralement stockées sur des bases de données (relationnelles ou no)). Cette recherche est effectuée à l'aide des méthodes mathématiques, statistiques ou algorithmiques. Data Mining se considère comme un processus le plus automatique possible, qui part de données élémentaires disponibles dans un Data Warehouse à la décision. L'objectif principale de Dat Mining c'est de créer un processus automatique qui a comme point de départ les données y comme finalité l'aide à la prise des décisions. (Voir Fig 2.1).

Data Mining versus KDD (Knowledge Discovery in Databases)

* habituellement les deux termes sont interchangeables.

- KDD (Knowledge Discovery in Databases) : C'est le processus de trouver information et/ou parts utiles à partir de données.
- Data Mining : C'est l'utilisation des algorithmes pour extraire information et/ou parts comme partie du processus KDD.

Data Mining: C'est une partie du processus KDD

Data Mining: Le coeur du processus d'extraction de connaissances.

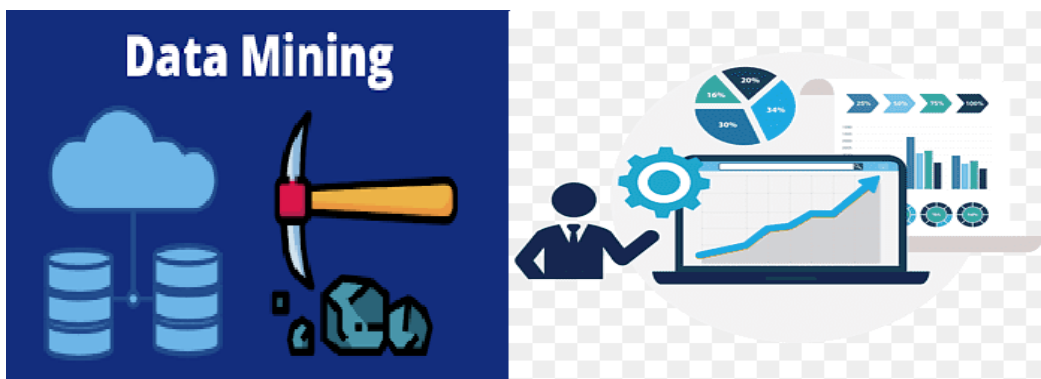


Figure 2.1: Logo Data Maining

1.1. L'utilité le data mining aujourd'hui:

Aujourd'hui, le data mining est utilisé dans de nombreux secteurs d'activité comme la recherche, le marketing, le développement de produits, la santé ou encore l'éducation.

Ce processus permet de résoudre rapidement des problèmes qui, jusqu'alors, demandaient énormément de temps pour être réglés manuellement.

L'utilisation de techniques statistiques diverses pour analyser les données permet aux utilisateurs d'identifier des modèles, des tendances et des corrélations qui n'apparaissent pas clairement au départ. Grâce aux résultats des différentes analyses successives, ils peuvent prédire ce qui est susceptible de se produire et prendre des mesures pour influencer et maximiser les résultats commerciaux.

Lorsque le data mining est employé efficacement, il peut fournir aux organisations un avantage considérable par rapport à leurs concurrents. Il permet en effet de mieux comprendre les clients, de développer des stratégies marketing efficaces, d'augmenter les revenus et de réduire les coûts.

1.2. Concepts clés du data mining:

Utiliser le data mining dans son entreprise implique de connaître de nombreux concepts, outils et techniques qui gravitent autour de cette notion. En voici les principaux :

- ◆ Le nettoyage et la préparation des données : c'est une étape au cours de laquelle les données sont transformées afin d'être analysées et traitées de manière opérationnelle. Supprimer des erreurs ou identifier une information manquante par exemple.
- ◆ L'intelligence artificielle (IA) : ce sont des systèmes qui effectuent des activités analytiques en imitant la réflexion humaine comme l'apprentissage, le raisonnement ou la résolution de problèmes.
- ◆ L'apprentissage de règles d'association : il s'agit d'outils recherchant des relations entre des variables d'un jeu de données. Il peut, entre autres, permettre à une entreprise d'identifier des produits qui sont généralement achetés ensemble par les clients.
- ◆ Le clustering : c'est un processus de partitionnement d'un jeu de données en clusters (sous-groupes) pour aider les utilisateurs à comprendre le regroupement de données ou des faits auparavant inconnus.
- ◆ La classification : cette technique sert à catégoriser ou classer des informations issues d'un jeu de données dans le but d'établir des prédictions.
- ◆ L'analyse de données : c'est le fait d'évaluer des informations numériques et de les utiliser de manière utile.
- ◆ Le data warehousing : il s'agit littéralement d'un entreposage de données afin d'aider une organisation à prendre les meilleures décisions. C'est une composante essentielle du data mining à grande échelle.

Le machine learning : c'est une technique de programmation informatique qui utilise des probabilités statistiques afin de donner aux ordinateurs et IoT la capacité « d'apprendre ». Le machine learning et l'intelligence artificielle sont deux notions liées.

La régression : c'est une méthode analytique utilisée pour prédire une plage de valeurs numériques (les ventes, les températures ou les cours des actions par exemple) à partir d'un jeu de données spécifique.

2.La modélisation prédictive:

La modélisation prédictive, ou Predictive Modeling en anglais, regroupe un ensemble de méthodes permettant de collecter et d'analyser des données définies, de manière à les interpréter pour en déduire des pronostics concernant des tendances futures, des événements à venir ou bien le comportement des consommateurs à l'avenir.

La modélisation prédictive donne ainsi lieu à des pronostics devant toutefois être considérés comme des probabilités et non comme des prédictions certaines, qui se concrétiseront forcément. La vraisemblance des résultats des modèles pronostics est à rapporter sur le plan statistique. Leur probabilité est alors envisageable en fonction de la taille de l'ensemble des données étudiées. Ainsi, plus le nombre de données analysées est important, plus les résultats des modèles de pronostics peuvent être considérés comme des résultats envisageables et précis. Il n'existe cependant aucune garantie quant à la survenue effective des données de sortie.

2.1. Les domaines d'application de La modélisation prédictive :

Plusieurs domaines et spécialités diverses ont régulièrement recours à la modélisation prédictive, tels que le secteur des assurances, le secteur financier, l'e-commerce, les télécommunications, la science, la gestion de la relation client ou encore l'informatique décisionnelle. La modélisation prédictive donne également lieu à des pronostics qui peuvent être utilisés comme bases décisionnelles permettant d'évaluer les chances et les risques dans le cadre d'une planification budgétaire.

La modélisation prédictive est ainsi couramment utilisée pour le calcul des risques dans le cadre d'une assurance-vie. Dans le domaine scientifique, cette science est employée pour confirmer ou infirmer des théories à l'aide de données issues d'un domaine de spécialité bien précis.

De manière générale, des pronostics peuvent être établis à l'aide d'un grand nombre d'analyses de régression et de modèles statistiques différents. Lorsque la modélisation prédictive est appliquée au domaine informatique, il s'agit notamment d'exploration de données (Data Mining) et d'apprentissage automatique (Machine Learning). Il convient alors d'extraire des données d'entrée pertinentes se présentant sous la forme de grands ensembles de données, et d'autre part, d'établir des modèles pronostiques capables d'auto-apprentissage, qui sont alors en capacité d'intégrer automatiquement de nouvelles données aux pronostics déjà établis.

Dans les domaines du web marketing, de l'analyse web, de l'optimisation du référencement et des réseaux sociaux, la modélisation prédictive permet également de calculer des revenus prévisionnels. Les pronostics établis servent ainsi à élaborer des planifications de budget, en effectuant des projections. Les secteurs web marketing peuvent ainsi cibler les canaux marketing les plus à même de toucher une cible prédéfinie. Cela est aussi valable en ce qui concerne la fidélisation des clients, la gestion de la relation client, la publicité ciblée sur les moteurs de recherche et réseaux affiliés, les envois de newsletters, etc. Les applications possibles de la modélisation prédictive sont ainsi très nombreuses.

2.2. Comment se déroule un processus type de modélisation prédictive?

Pour trouver les bonnes solutions aux problèmes opérationnels grâce à la modélisation prédictive, il faut collaborer de près avec les spécialistes en la matière du début à la fin. Voici les éléments centraux du processus (Voir Fig 2):

- 1- Cerner un problème pour lequel les prévisions des résultats ou du comportement futur peuvent faire en sorte de prendre des décisions d'affaires plus exactes et efficaces.
- 2- Comprendre l'activité : connaître les produits, les besoins des intervenants, les ressources disponibles et les données dont on dispose, et déterminer les hypothèses et les contraintes ainsi que les moyens de mettre en place une solution de modélisation prédictive.

- 3- Définir clairement le résultat que doit prédire (la variable de réponse) le modèle prédictif.

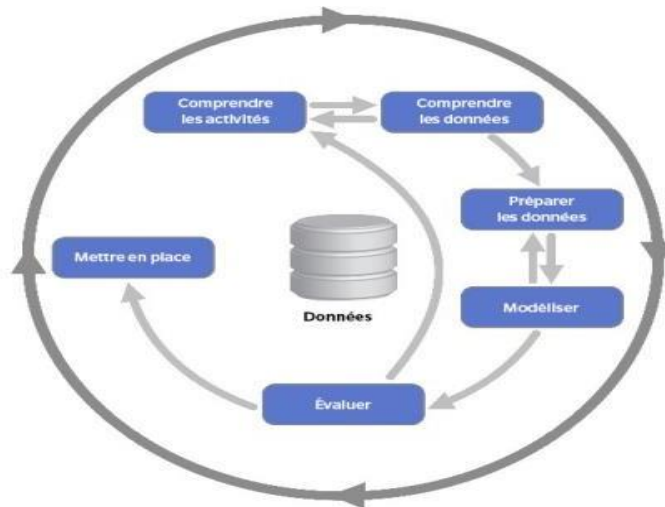


Figure 2.2: La modélisation prédictive.

2.3. Les modèles d'analyse prédictive

les modèles prédictifs n'ont pas besoin d'être créés à partir de zéro pour chaque application. Les outils d'analyse prédictive utilisent une variété de modèles et d'algorithmes approuvés qui peuvent être appliqués à un large éventail de cas d'utilisation. Les techniques de modélisation prédictive ont été perfectionnées au fil du temps. Au fur et à mesure que nous ajoutons plus de données, plus d'informatique musculaire, d'IA et d'apprentissage automatique et que nous constatons des progrès globaux dans l'analyse, nous sommes en mesure d'en faire plus avec ces modèles.

Les cinq principaux modèles d'analyse prédictive sont :

2.3.1. Modèle de classification :

Le modèle de classification est, à certains égards, le plus simple des différents types de modèles d'analyse prédictive que nous allons couvrir. Il met les données dans des catégories en fonction de ce qu'il apprend des données historiques.

Les modèles de classification sont les meilleurs pour répondre oui ou non aux questions, fournissant une analyse large qui est utile pour guider une action décisive.

L'étendue des possibilités offertes par le modèle de classification - et la facilité avec laquelle il peut être reconverti avec de nouvelles données - signifie qu'il peut être appliqué à de nombreuses industries différentes.

➤ **Classification :**

La classification est le processus d'identification et de regroupement d'objets ou d'idées en catégories prédéterminées. Dans la gestion des données, la classification permet la séparation et le tri des données en fonction des exigences définies pour divers objectifs commerciaux ou personnels. Dans l'apprentissage automatique (ML), la classification est utilisée dans la modélisation prédictive pour attribuer des données d'entrée avec une étiquette de classe. Par exemple, un programme de sécurité de courrier électronique chargé d'identifier le spam pourrait utiliser traitement du langage naturel (PNL) pour classer les courriels comme étant « spam » ou « non spam ».

➤ **Prédiction :**

En prédiction, on distingue les problèmes de régression des problèmes de classement. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression tandis que les problèmes de prédiction d'une variable qualitative sont des problèmes de classification.

1-Construction du modèle :

Chaque objet appartient à une classe connue. Le jeu de données d'apprentissage est l'ensemble des objets utilisés pour la construction du modèle.

1-Utilisation du modèle pour classer des objets :

Estimation de la précision du modèle : les classes connues du jeu d'apprentissage sont comparées à celles prédites. Le pourcentage d'objets de jeu de test correctement classés.

Le jeu de test est indépendant du jeu d'apprentissage sinon risque de biais et de sur-apprentissage. (Voir Fig 2.3)

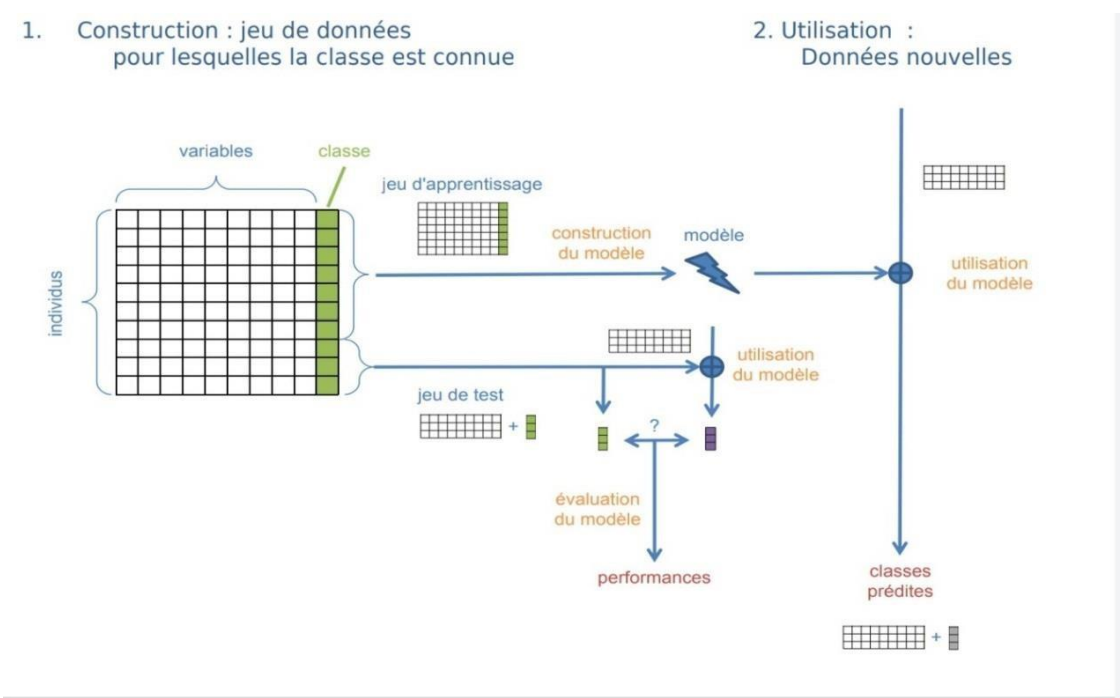


Figure2. 3 : Schéma de model de classification

1-Classification : Construction du modèle(Voir Fig 4).

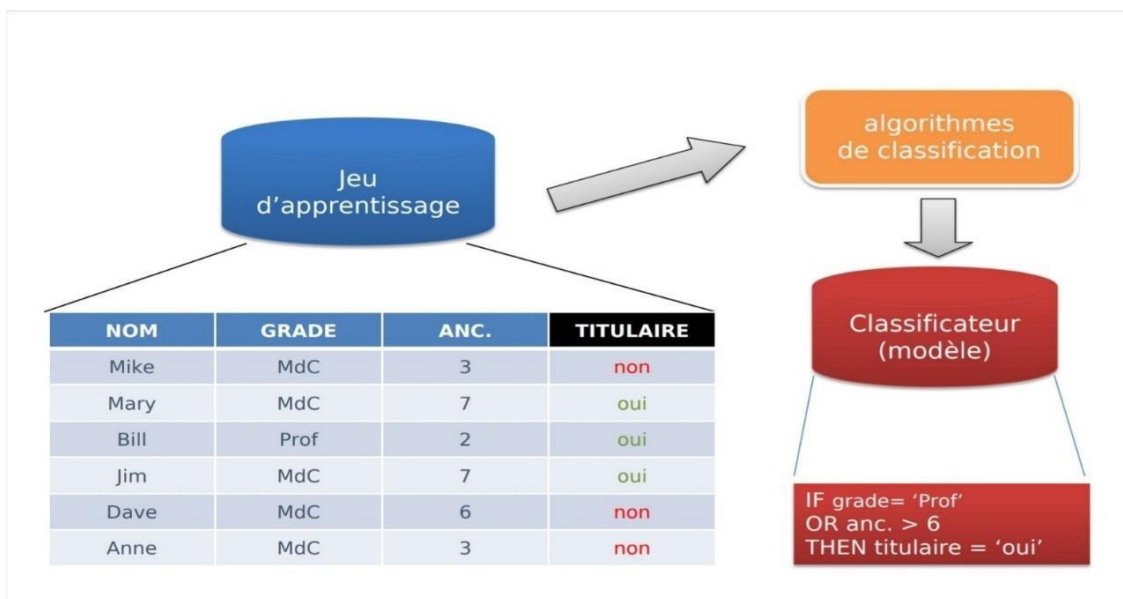


Figure 2.4: Construction du modèle

2-Classification : utilisation du modèle (Voir Fig 2.5).

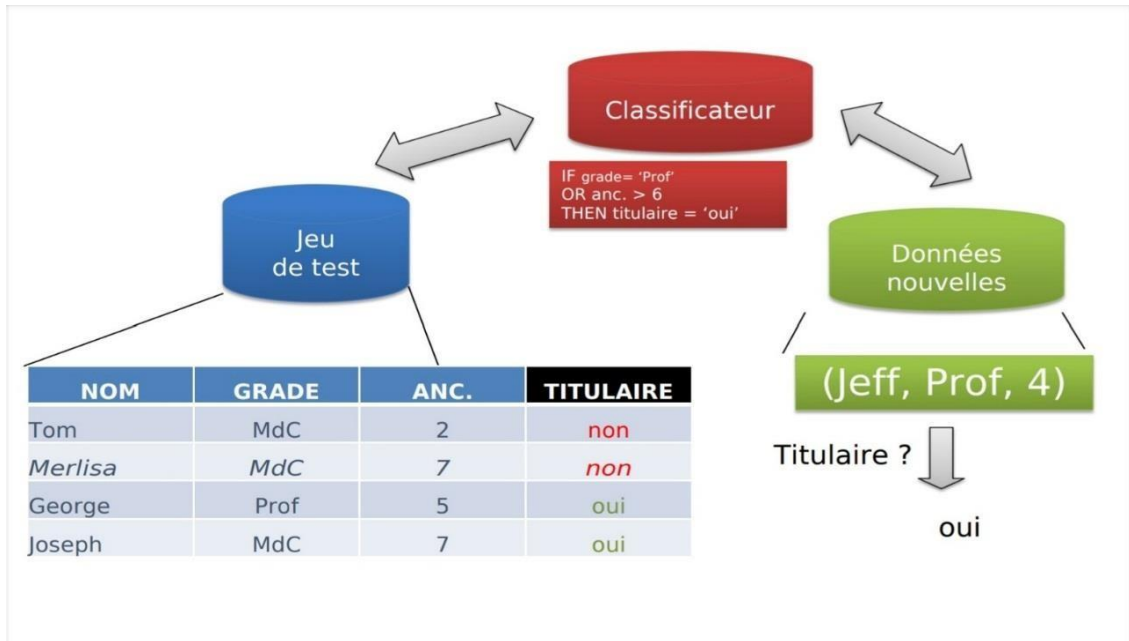


Figure 2.5: Utilisation du modèle

• **Méthodes :**

- Arbre de décision et forêt aléatoire
 - classificateur bayésien (naïf ou réseau bayésien)
 - Réseau de neurones (perceptron et deep learning)
- Machine à vecteur de support

2.3.1.1-Classification par arbre de décision :

Un arbre de décision DT (Decision Tree) est un enchaînement hiérarchique de règles logiques ou de production construite de manière automatique à partir d'un ensemble de données. La construction de l'arbre de décision consiste à utiliser les descripteurs, pour subdiviser progressivement l'ensemble en sous-ensembles de plus en plus fins. L'algorithme estime la probabilité qu'une observation se trouve dans le nœud t en utilisant la formule :

$$P(T) = \sum_{j \in T} w_j$$

Où

w_j : Poids des observations,

T : Ensemble de tous les indices d'observation du nœud

n : Égale au nombre d'observations.

Les arbres de décision sont des algorithmes de classification de données reposant sur

un apprentissage supervisé. Comme son nom l'indique, la structure de ce algorithmes ressemble à des arbres constitués de nœuds, de branches et de feuilles (voir figure 2.4). La construction de ces arbres est réalisée à l'aide d'une base de données brute (vecteur des caractéristiques et classes) et de lois qui permettent de déterminer les variables discriminantes pour la classification efficace des futures données. Chacun des nœuds constituant l'arbre représente une règle de classification préalablement déterminée de manière récursive. (Voir Fig 2.6).

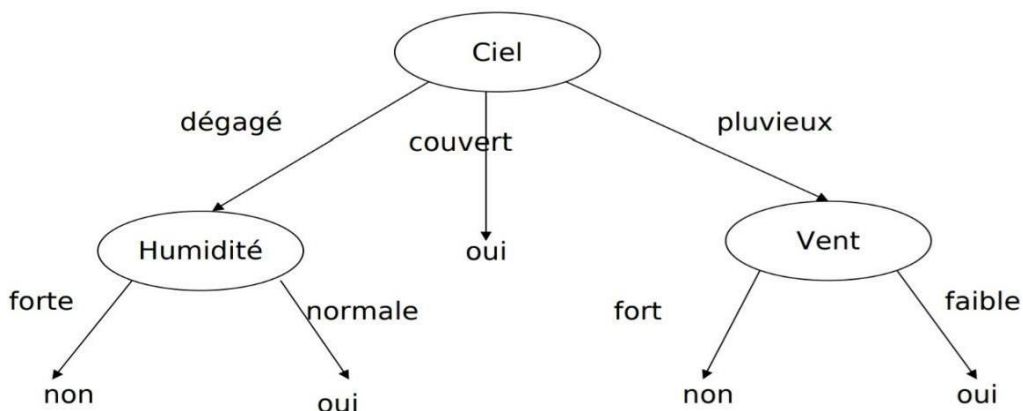


Figure 2.6: L'arbre de décision

Exemple : pratique la méditation(Voir Fig 2.7).

âge	stress	étudiant	alimentation	médite
<=30	élevé	non	bonne	non
<=30	élevé	non	excellente	non
31...40	élevé	non	bonne	oui
>40	moyen	non	bonne	oui
>40	bas	oui	bonne	oui
>40	bas	oui	excellente	non
31...40	bas	oui	excellente	oui
<=30	moyen	non	bonne	non
<=30	bas	oui	bonne	oui
>40	moyen	oui	bonne	oui
<=30	moyen	oui	excellente	oui
31...40	moyen	non	excellente	oui
31...40	élevé	oui	bonne	oui
>40	moyen	non	excellente	non

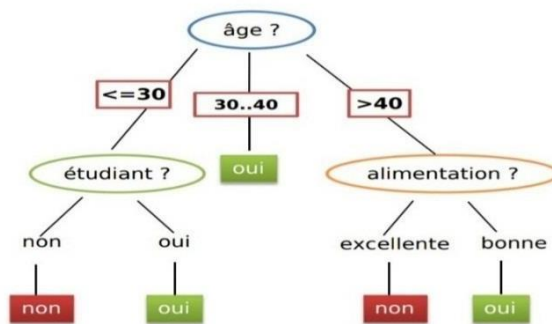


Figure 2.7: Exemple de l'arbre de décision

2.3.1.2. Classification Bayésienne :

Apprentissage probabiliste : calcule explicitement les probabilités des hypothèses, une des approches les plus pragmatiques pour certains types d'apprentissage.

Incrémental : chaque exemple met à jour la probabilité qu'une hypothèse est correcte. Des connaissances a priori peuvent être combinées avec des données d'observation.

Prédiction probabiliste : prédit plusieurs hypothèses, pondérées par leur probabilité

Le problème de classification peut être formulé en utilisant

les probabilités a posteriori : $P(C|X)$ = probabilité que l'objet $X = \langle x_1, \dots, x_k \rangle$ est de la classe C

Théorème de Bayes:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

$P(\text{classe} | \text{observations}) = P(\text{observations} | \text{classe}) \cdot P(\text{classe}) / P(\text{observations})$

Principe : attribuer la classe la plus probable

C_i tel que $P(C_i|X)$ est maximum

C_i tel que $P(X|C_i) \cdot P(C_i) / P(X)$ est maximum

remarque : $P(X)$ est constant pour toutes les classes donc pas besoin de la calculer pour trouver la classe qui aura la plus forte probabilité

$P(C)$ = fréquence de la classe C .

2.3.1.3. Machines à Vecteurs de Support (SVM) :

Les machines à vecteurs supports (Support Vector Machines, SVM) sont des méthodes généralistes d'apprentissage et de discrimination. Ces méthodes ont montré leur efficacité dans de nombreuses applications.

Les SVM ont été à l'origine des travaux de Vapnik et ont été conçus pour la décision binaire. L'originalité principale de ces méthodes consiste à utiliser efficacement les exemples étiquetés afin de produire une fonction de décision qui maximise la marge entre deux classes données. L'efficacité repose sur la sélection des éléments d'apprentissage les plus représentatifs de la tâche de décision tout en maximisant la capacité de généralisation du modèle. Ces éléments correspondent aux exemples proches de la frontière de décision. Ils sont appelés les supports vecteurs et sont situés sur la marge.

Pour garantir la bonne généralisation de la fonction de décision recherchée, le problème mathématique correspond à la maximisation de la distance des exemples annotés de la frontière (Voir Fig. 2.8).

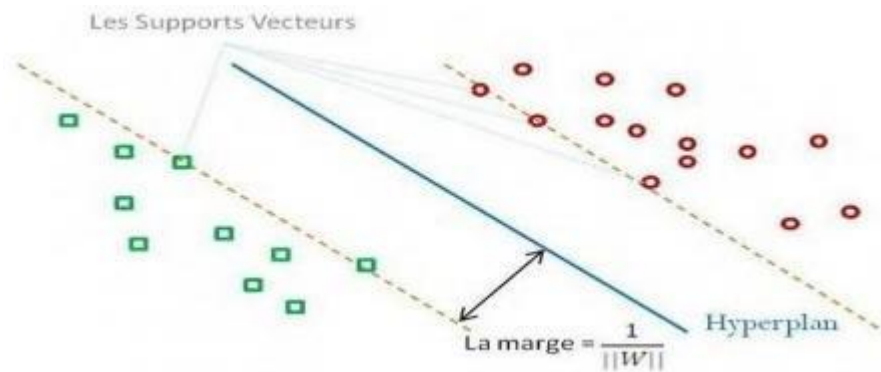


Figure 2.8: Méthode à noyau SVM.

2.3.2. Modèles de Régression :

La régression est une technique statistique utilisée dans divers domaines pour identifier la force et la nature d'un lien entre une variable dépendante (généralement indiquée par Y) et un ensemble d'autres variables appelées variables indépendantes. Tout attribut à valeur continue peut-être prédit à l'aide de la régression, qui est une forme d'approche d'apprentissage automatique supervisée. Toute organisation commerciale peut utiliser la régression pour examiner les corrélations entre la variable cible et la variable prédictive.

(Voir Fig. 2.9).

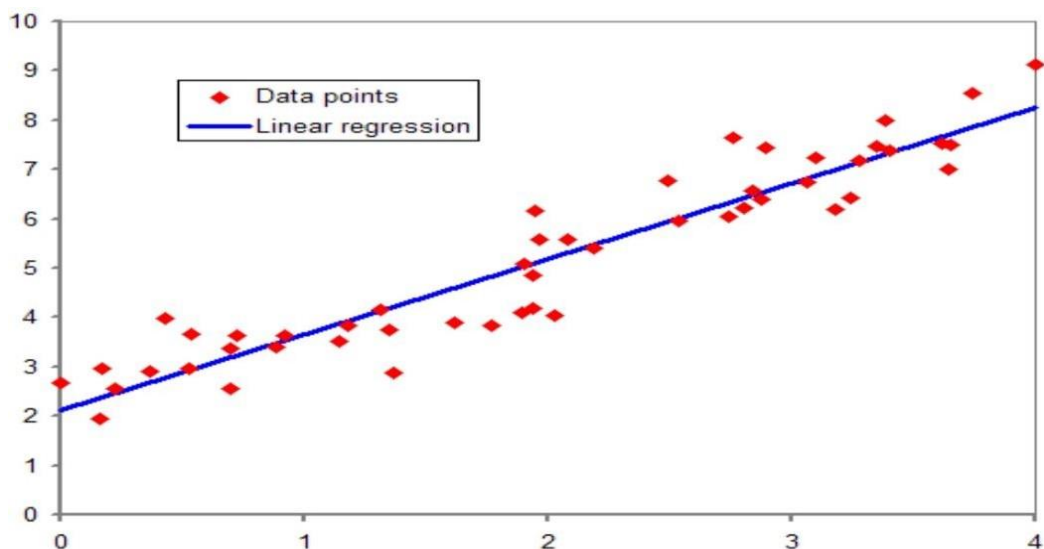


Figure 2.9: Modèle de Régression.

2.3.2.1. Régression linéaire

La régression linéaire est l'une des plus anciennes techniques de prédiction en statistique. L'objectif de la régression linéaire est d'adapter un modèle linéaire entre la réponse et les variables indépendantes, et de l'utiliser pour prédire le résultat compte tenu d'un ensemble de variables indépendantes observées. Un modèle de régression linéaire simple utilise une formule de la forme :

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \dots + \varepsilon \quad (2.1)$$

Où :

- Y est la variable de réponse (c'est-à-dire le résultat qu'on essaie de prédire).
- X1, X2, X3, etc. sont les variables indépendantes utilisées pour prédire le résultat.
- B0 est une constante qui est l'intersection de la droite de régression.
- B1, b2, b3, etc. sont les coefficients des variables indépendantes.

Il s'agit des pentes partielles de chaque variable. - ε est l'erreur ou le bruit associé à la variable de réponse qui ne peut pas être expliqué par les variables indépendantes X1, X2, et X3. Un modèle de régression linéaire a deux composantes : une partie déterministe (i.e. $b_1X_1 + b_2X_2 + \dots$) et une partie aléatoire (i.e. l'erreur, ε). On peut considérer ces deux composants comme le signal et le bruit dans le modèle. Si On n'a qu'une seule variable d'entrée X, le modèle de régression est la meilleure ligne qui correspond aux données.

La figure 10 montre un exemple de modèle de régression linéaire simple. Avec deux variables d'entrée, la régression linéaire est le meilleur plan qui s'adapte à un ensemble de points de données dans un espace 3D. Les coefficients des variables (i.e. b_1, b_2, b_3 , etc.) sont les pentes partielles de chaque variable. Si on maintient toutes les autres variables constantes, le résultat Y augmentera de b_1 lorsque la variable X1 augmentera de 1. C'est pourquoi les économistes utilisent généralement l'expression « ceteris partibus » ou « toutes choses étant égales par ailleurs » pour décrire l'effet d'une variable indépendante sur un résultat donné. (Voir Fig. 2.10).

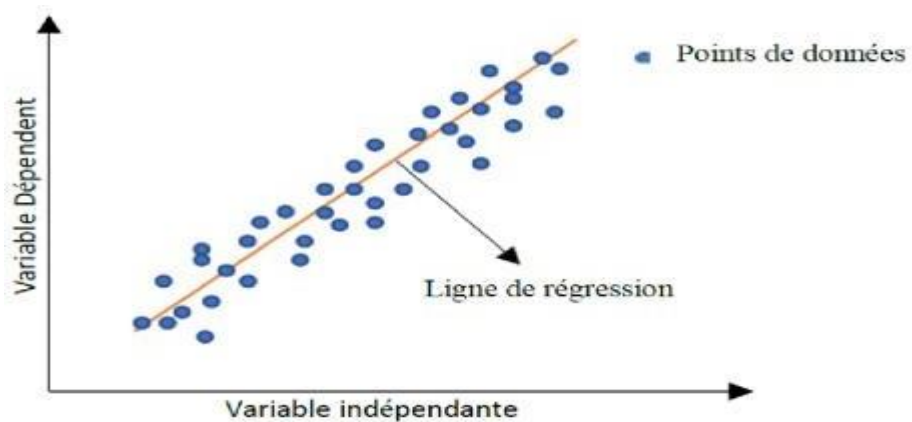


Figure 2.10: Une relation linéaire simple entre deux variables.

La régression linéaire utilise les méthodes des moindres carrés ou de descente de gradient pour trouver les meilleurs coefficients de modèle pour un ensemble de données fournies. La méthode des moindres carrés y parvient en minimisant la somme de l'erreur quadratique entre les valeurs ajustées et réelles de chaque observation dans les données d'apprentissage.

La descente de gradient trouve les coefficients de modèle optimaux en mettant à jour les coefficients à chaque itération. Les mises à jour vont dans le sens où la somme des erreurs entre le modèle ajusté et les valeurs réelles des données d'apprentissage est réduite. Par plusieurs itérations, il trouve le minimum local en se déplaçant dans le sens du gradient négatif.

3-Quelques meilleurs outils pour faire en data méninge :

3.1-Python :

Python est un langage de programmation très puissant utilisé en Data Mining pour faire de l'analyse statistique, la classification, le clustering et l'analyse prédictive. (Voir Fig. 2.11).



Figure 2.11: Logo python

3.2-Le langage R

R est un langage de programmation et un logiciel libre destiné aux statistiques et à la science des données soutenu par la R Foundation for Statistical Computing. Il permet de faire l'analyse statistique, la classification, le clustering et l'analyse prédictive. (Voir Fig. 12).

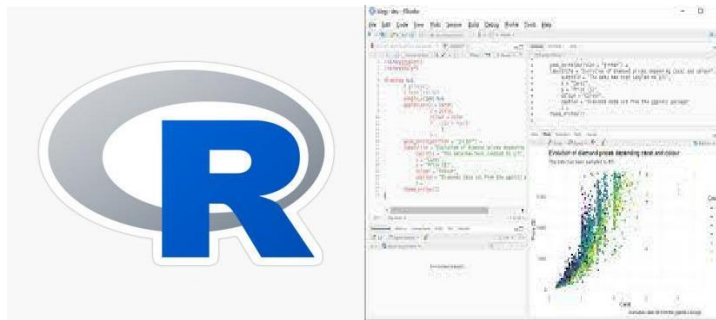


Figure 2.12: Logo R.

3.3-Tanagra

Tanagra est un logiciel gratuit de Data Mining destiné à l'enseignement et à la recherche. Il implémente une série de méthodes de fouilles de données issues du domaine de la statistique exploratoire, de l'analyse de données, de l'apprentissage automatique et des bases de données. C'est un projet ouvert au sens qu'il est possible à tout chercheur d'accéder au code et d'ajouter ses propres algorithmes pour peu qu'il respecte la licence de distribution du logiciel. (Voir Fig. 13).

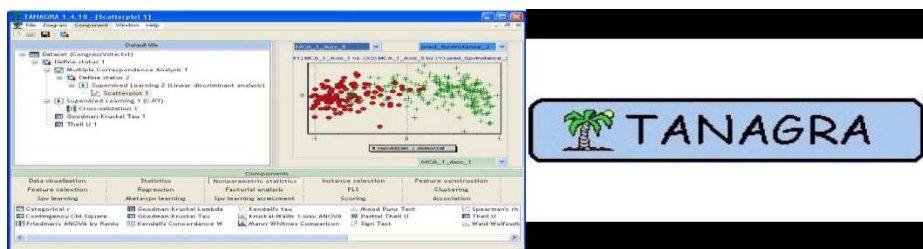


Figure 2.13 : Logo Tanagra.

Conclusion :

Dans ce chapitre, nous avons présenté le DM et certains de ses algorithmes. Nous avons d'abord commencé par décrire le machine Learning et ses différentes architectures proposées. Ensuite, nous sommes passés à une description détaillée des algorithmes d'apprentissage supervisé.

CHAPITRE 3 Etat de l'art

Introduction :

Le comportement des touristes a des implications importantes pour la gestion et le marketing des destinations. Comprendre le comportement des touristes aide les destinations à développer des stratégies efficaces pour attirer et retenir les touristes. Il aide également les entreprises à adapter leurs produits et services pour répondre aux besoins et aux préférences des touristes, divers systèmes basés sur le DM ont été mis au point pour améliorer la classification et la prédiction du comportement des touristes, et ils dépendent sur des données étiquetées.

Daniel Zeppelin et al. Ont utilisé un réseau d'attention à plusieurs niveaux pour calculer la relation spatiale temporelle entre les sites non contigus et les visites non contiguës à la séquence de parcours de l'utilisateur, et mettre à jour la représentation de chaque visite qui confère au chemin de l'utilisateur. Les résultats expérimentaux sur deux ensembles de données du monde réel (New York et Guala City) montrent qu'ils sont efficaces et largement supérieurs aux méthodes existantes.

Tingui Chen. A proposé une méthode pour extraire des modèles de comportement touristique basés sur un diagramme de réseau. L'auteur a créé d'abord cette méthode (a) diagramme de réseau. Ensuite, il a obtenu des modèles de comportement très intéressants pour extraire les règles de corrélation dans les modèles et obtenir les règles. Enfin, il a expérimenté en montrant la faisabilité et l'efficacité de la méthode dans le lac de l'Ouest de Hangzhou

G. Gallo et al. ont utilisé des techniques d'exploration de données (data mining techniques), en utilisant des images géographiques accessibles au public sur les plateformes de médias sociaux comme source d'information pour mieux comprendre le comportement des touristes, et ont également suggéré un moyen de déduire le flux de touristes dans la région autour de l'Etna.

Shah J Miah et al. ont utilisé la méthode d'analyse de données volumineuses pour analyser le comportement du tourisme à l'aide de la méthodologie DSR pour extraire, classer, localiser et quantifier des informations touristiques significatives à partir de grands ensembles de données non structurées afin de soutenir la prise de décision stratégique du DMO, comme le montre le cas de Melbourne, Australie.

1.Travail 01: Next Location Recommendation with Attention Mechanism by Mining the Spatio-Temporal Relationship between Visited Locations

Les réseaux sociaux géolocalisés (LBSN) permettent aux utilisateurs de socialiser avec leurs amis en partageant leurs expériences quotidiennes en ligne. En particulier, une grande quantité de données de check-ins générées par les utilisateurs capture les lieux de visite des utilisateurs et ouvre une nouvelle ligne de recherche de mégadonnées spatio-temporelles, c.-à-d. la recommandation suivante sur le point d'intérêt (PI). À l'heure actuelle, certaines méthodes avancées ont été proposées pour la recommandation de l'IPE, les travaux en cours tirent seulement parti des renseignements temporels de deux vérifications consécutives du LBSN. Plus précisément, ces méthodes ne visent que les visites adjacentes, mais ignorent les visites non contiguës, alors que ces visites peuvent être importantes pour comprendre la corrélation spatio-temporelle au sein de la trajectoire.

Afin d'extraire l'information de cette visite non contiguë, on propose un modèle multi-couches dans l'apprentissage temporel spatial profond Time Spatial Converter Recommendation (STTF-Recommen).

Le modèle proposé a été comparé aux modèles de base, où il a été jugé obtenir de meilleurs résultats que les autres modèles de référence. Chaque modèle fonctionne 10 fois avec différents ensembles de données et différents pics, ils ont utilisé la performance moyenne de chaque modèle d'évaluation. Les résultats indiquent que le modèle proposé a été amélioré d'au moins 4 % par rapport aux meilleurs modèles de référence, ce qui démontre la faisabilité et l'efficacité du modèle.

2.Travail 02:Tourist Behavior Pattern Mining Model Based on Context

L'expérience de voyage personnalisée et le service de tourisme a été un sujet brûlant de recherche dans la chaîne d'approvisionnement des services touristiques. Dans ce document, nous prenons le contexte en considération et proposons une méthode analysée au touriste sur la base du contexte : premièrement, nous analysons le contexte qui influence les modèles de comportement touristique, sélectionner les principaux facteurs de contexte, et construire le modèle de comportement touristique; puis, nous calculons le degré d'intérêt du modèle de comportement touristique et de mine sur les règles avec un grand intérêt degré avec l'algorithme de règle d'association; nous pouvons faire quelques recommandations au touriste avec un voyage mieux personnalisé expérience et services.

La méthode que nous proposons dans cet article, nous utilisons l'algorithme Apriori, l'algorithme GRI, l'algorithme CARAMA sur les données originales (ici nous fixons le support minimum à 20% et le minimum confiance égale à 80%). Dans cette expérience la validité de la méthode est d'environ 91,67% (11/12) à l'Apriori et 100% (6/6) à l'algorithme GRI. Cela signifie que la méthode que nous avons proposé est réalisable et efficace.

Nous avons considéré les facteurs de contexte qui influencent le modèle de comportement touristique globalement, tels que l'appareil l'utilisation touristique, l'heure, l'emplacement, et les types de service, et avons obtenu l'ensemble de contexte qui influence le modèle de comportement touristique. Ensuite, nous avons proposé une méthode pour exploiter les modèles de comportement touristique basé sur le diagramme de réseau; cette méthode construit un diagramme de réseau d'abord. Ensuite, nous avons obtenu les modèles de comportement avec un degré intéressant élevé et fait l'exploitation des règles d'association dans les modèles et a obtenu les règles; enfin, nous avons fait une expérience pour montrer la faisabilité et efficacité de notre méthode.

3.Travail 03: Exploiting Social Images to Understand TouristBehaviour.

Dans cet article, il suggère d'exploiter les images géographiques accessible au public sur les plateformes de médias sociaux comme source d'information pour un aperçu du comportement touristique. Au départ, les métadonnées d'images géographiques sont

explorées à l'aide d'outils d'analyse visuelle pour cerner les tendances, les tendances et les relations entre l'information obtenue par les médias sociaux. Ensuite, des techniques d'extraction de données sont utilisées pour générer un modèle d'itinérance pour la zone d'intérêt. Enfin, nous examinons les sites qui peuvent être visités conjointement et analysons la façon dont le tourisme se déplace dans ces pour confirmer l'efficacité de l'analyse proposée, nous avons testé les méthodes proposées dans une étude de cas.

Diverses méthodes utiles ont été proposées pour analyser le flux touristique dans une zone importante et pour extraire des guides utiles qui aident les gestionnaires du tourisme à mener une activité de planification touristique fiable. En particulier, nous avons fondé notre analyse sur les données disponibles aujourd'hui sur les plateformes de médias sociaux. Nous avons construit un outil pour analyser la répartition spatiale des touristes. En utilisant des techniques d'exploration de données, nous avons également suggéré une façon de déduire le flux de touristes dans la zone d'intérêt. Cette méthode a été appliquée à une étude de cas spécifique, en tenant compte de la zone autour de l'Etna, l'un des sites sur la Liste du patrimoine mondial. Les résultats appuient et confirment les connaissances des experts sur cette région.

4.Travail 04: A big data analytics method for tourist behavior analysis

Les Big data générées sur les sites de médias sociaux ont créé de nombreuses occasions d'apporter plus de renseignements aux décideurs. Cependant, peu d'études sur l'analyse des mégadonnées ont démontré le soutien à la prise de décisions stratégiques. En outre, une méthode formelle d'analyse des mégadonnées générées par les médias sociaux pour l'aide à la décision n'a pas encore été développée, en particulier dans le secteur du tourisme.

À l'aide d'une approche de recherche en sciences de la conception, cette étude vise à concevoir et à évaluer une méthode d'analyse des « Big data » pour appuyer la prise de décisions stratégiques dans la gestion des destinations touristiques. Utilisation de photos géolocalisées téléchargées par les touristes sur le site de partage de photos sur les médias sociaux, Flickr, l'applicabilité de la méthode pour aider les organisations de gestion de destination à analyser et à prédire les modèles de comportement des touristes à des endroits précis destinations est montré, en utilisant Melbourne, Australie, comme un cas représentatif. Le service public a été confirmé en utilisant à la fois une autre destination et directement auprès des intervenants. L'artefact développé démontre une méthode d'analyse

des mégadonnées non structurées pour améliorer la prise de décision stratégique dans un domaine à problèmes réel. La méthode proposée est générique et son applicabilité à d'autres flux de mégadonnées est discutée.

Nous avons fourni un moyen d'extraire, de classer, d'identifier et d'identifier les informations touristiques pertinentes Ensembles de mégadonnées non organisés pour appuyer la prise de décisions stratégiques du BSM. Analyser les données géographiques Avec d'autres détails pertinents, notre méthode s'applique à différentes destinations et est créée Résultats utiles, comme indiqué dans le cas de Melbourne, Australie. Nous avons suivi la théorie établie et Lignes directrices méthodologiques du Comité pour la conception, le développement et la diffusion des produits artefact ; Une façon, l'un des quatre types d'artefacts de concepteur reconnus dans la littérature DSR dans l'information (). Nous avons utilisé MATLAB, environnement informatique numérique, avec Google Mas, un Service de cartographie Web de bureau, comme plate- forme/environnement technique pour développer et évaluer la solution (). Pour d'autres améliorations techniques à notre méthode d'analyse proposée, algorithmes avancés connexes améliorer les capacités d'analyse.

Conclusion :

Dans ce chapitre, ils présentent quelques travaux antérieurs sur le sujet de l'étude.

Auteurs	Titre de l'article	Lieu d'étude	Approche utilisée
Danielle Zibeline et al	Next Location Recommendation with Attention Mechanism by Mining the Spatio-Temporal Relationship between Visited Locations	New York City and Gowalla	The STTF- Recommender Model
Tinggui Chen	Tourist Behavior Pattern Mining Model Based on Context	the West Lake of Hangzhou	Network diagram model
G. Gallo et al	Exploiting Social Images to Understand Tourist Behaviour	The mount Etna	data mining techniques
Shah J Miah et al	A big data analytics method for tourist behaviour analysis	the case of Melbourne, Australia	DSR methodology

Tableau 3.1: Etat de l'art

CHAPITRE 4 : Approche proposée, Résultats et Discussion

Introduction :

La compréhension du comportement touristique aide les destinations à élaborer des stratégies efficaces pour attirer et retenir les touristes. Il aide également les entreprises à adapter leurs produits et services aux besoins et préférences des touristes, et divers systèmes de gestion des sinistres ont été développés pour améliorer la classification et la prédiction du comportement touristique, et ils s'appuient sur des données étiquetées.

Dans ce chapitre, nous présenterons notre modèle proposé pour prédire le comportement touristique à l'aide de l'apprentissage automatique supervisé avec classification et régression.

1.Lieu d'étude : (wilaya de M'sila)

1.1. Définition

M'sila, la « capitale de la couvée » est située au centre du nord algérien. C'est un lien entre la colline et le désert. Elle appartient à la région des hauts plateaux, avec une superficie de 18 175 km², et une population de plus d'un million et deux cent mille personnes. Elle est devenue un État selon la division administrative de 1974, composé de 47 municipalités et 15 départements.

L'Etat de M'sila se caractérise également par un long passé, une histoire glorieuse, et une situation stratégique importante et unique, ces atouts touristiques de notre État le qualifient pour être un pôle touristique qui joue un rôle important dans le développement local et national en contribuant à l'emploi, à la création de postes de travail, notamment pour les jeunes et à la création de richesses. Ce secteur stratégique a besoin d'être soutenu et accompagné dans Afin de jouer son rôle effectif dans la société et l'économie nationale, Le secteur du tourisme et de l'artisanat est connu pour employer un grand nombre de main-d'œuvre, car c'est l'un des secteurs fiables dans la création d'emplois permanents, temporaires et saisonniers.

1.2. Zones et sites touristiques

- Qalat bani hammed
- Zawiya el hamel
- Chalalt karma
- Chott el hodna
- La restaurant blanc

2-Méthodologie et préparation des données

Nous avons obtenu des données sur les statistiques d'arrivées et de nuitées de différentes nationalités algériennes et étrangères à partir d'extraits de rapports préparés par la Direction du Tourisme de l'État de M'sila. Ces données s'étendaient sur une période de 2012 jusqu'à 2022 (Voir Fig. 4.1).

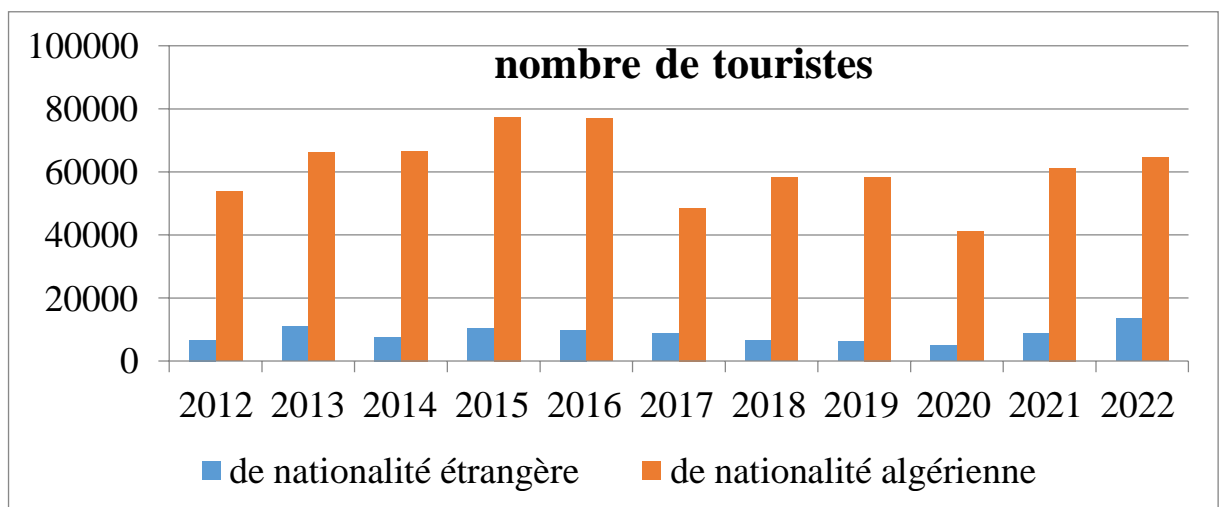


Figure 4.1: Nombre de touristes dans M'sila (2012-2022)

Nous avons enregistré les données dans des tableaux et intégrées dans un fichier Excel. Ces données peuvent inclure les variables suivantes : pays, Age, le sexe, Budget, Saison, type de site , Destination. (Voir Fig. 4.2)

L'ensemble de données :

pays : le pays d'origine du touriste (par exemple, Algérie, France, Italy,.....)

Age : l'âge du touriste (numérique : 15, 25, 40, ...)

le sexe : le sexe du touriste (catégoriel : Homme, Femme)

Budget : la fourchette de budget du touriste pour le voyage (catégoriel : faible , moyen, fort).

Saison : la saison ou la période de l'année préférée pour la visite (catégoriel : par exemple , été, hiver, printemps, automne).

type de site : type de destination touristique (catégoriel : par exemple historique, loisirs,.....)

Destination : la destination touristique prévue (catégoriel : par exemple , qalat bani hammed , zawiya el hamel,.....)

Num	pays	Age	le sexe	Budget	Saison	type de site	Destination
1	Algérie	45	femme	fort	Printanière	Historique	Qalat Bani Hammad
2	France	20	femme	faible	Printanière	Religieux	zawiya el hamel
3	Maroc	17	femme	faible	Hiverner	loisirs	Qalat Bani Hammad
4	Tunisie	42	homme	fort	Automne	loisirs	chalalt kamra
5	Italy	27	homme	Moyen	Automne	Religieux	zawiya el hamel
6	Algérie	30	homme	Moyen	Printanière	Historique	Qalat Bani Hammad
7	canada	32	femme	Moyen	Automne	Historique	Qalat Bani Hammad
8	Espagne	31	homme	Moyen	Hiverner	Religieux	Qalat Bani Hammad
9	Egypte	40	femme	fort	Hiverner	autre	zawiya el hamel
10	Turquie	43	femme	fort	Hiverner	Historique	Qalat Bani Hammad
11	France	48	homme	fort	Automne	loisirs	chalalt kamra
12	Allemagne	29	homme	Moyen	Printanière	loisirs	chalalt kamra
13	Libye	21	homme	Faible	Printanière	Religieux	zawiya el hamel
14	canada	32	homme	Moyen	Automne	autre	zawiya el hamel
15	Espagne	48	homme	fort	Hiverner	Historique	Qalat Bani Hammad
16	Algérie	21	homme	Faible	Hiverner	loisirs	chalalt kamra
17	Algérie	34	homme	Moyen	Hiverner	Historique	Qalat Bani Hammad
18	France	18	homme	Faible	Automne	Religieux	zawiya el hamel
19	Italy	42	homme	fort	Automne	Historique	Qalat Bani Hammad
20	Allemagne	17	femme	Faible	Printanière	Historique	Qalat Bani Hammad
21	Italy	39	homme	Moyen	Hiverner	loisirs	chalalt kamra
22	Algérie	15	homme	Faible	Hiverner	loisirs	chalalt kamra

Figure 4.2 : Extrait de la Base de données en Excel

3- Les outils et détails d'implémentation

3.1-Langage de programmation

Au cours des dernières années, de nombreux langages de programmation ont été disponibles, Parmi les langages, notre sélection s'est concentrée sur Python.

3.1.1-Python

Python est un langage stable et flexible qui fournit aux développeurs une variété d'outils. Quelle Faites-en un choix de premier ordre pour le DM, du développement, la mise en œuvre et l'entretien. Python aide les développeurs à développer des produits Confiance [30].

Python présente de nombreux avantages :

- Simple et cohérent : la simplicité de Python facilite la gestion des développeurs Algorithmes complexes pour le ML.
- Flexibilité : Le facteur de flexibilité réduit la possibilité d'erreur, il laisse Le programmeur a le contrôle total et s'y engage Confortable.
- Bibliothèques et Framework : Algorithmes de Machine Learning Très complexe, mais Python est une solution de repli avec une large gamme Bibliothèques telles que Scikit-learn, Keras et Framework TensorFlow.
- Lisibilité : Python est facile à lire, les développeurs peuvent donc Il est facile de comprendre le code.
- Indépendance de la plate-forme : Python est un langage indépendant de la plate-forme. Plate-forme. Il est pris en charge par de nombreuses plates-formes, notamment Windows, Linux et MacOS.

3.1.2-Bibliothèques Python

Sklearn:Scikit-learn (sklearn) est une bibliothèque Python gratuite pour le ML. Il est développé par de nombreux contributeurs², notamment universitaires, par des instituts d'enseignement supérieur et de recherche français comme Inria³. Il fournit de nombreuses bibliothèques d'algorithmes dans son Framework à implémenter, clé en main.

Ces bibliothèques sont particulièrement utiles pour les data scientists. Il comprend spécifiquement des fonctions d'estimation de forêts aléatoires, de régression logistique, d'algorithmes de classification et de machines à vecteurs de support. Il est conçu pour se coordonner avec d'autres bibliothèques Python gratuites, notamment NumPy et SciPy

Pandas :Pandas est une bibliothèque open source conçue principalement pour travailler avec des données relationnelles ou étiquetées de manière simple et intuitive. Il fournit diverses structures de données et opérations pour manipuler des données numériques et des séries chronologiques. Cette bibliothèque est construite au-dessus de la bibliothèque NumPy. Pandas est rapide et offre des performances et une productivité élevée pour les utilisateurs.

NumPy : est une bibliothèque Python utilisée pour travailler avec des tableaux. Elle possède également des fonctions pour travailler dans le domaine de l'algèbre linéaire, de la transformée de Fourier et des matrices.

Matplotlib :est une bibliothèque multiplateforme de visualisation de données et de traçage graphique pour Python et son extension numérique NumPy. En tant que tel, il offre une alternative open source viable à MATLAB.

3.2. Environnement de développement

3.2.1. Jupyter

Jupyter est une application web utilisée pour programmer dans plus de 40 langages de programmation, dont Python, Julia, Ruby, R, ou encore Scala. C'est un projet communautaire dont l'objectif est de développer des logiciels libres, des formats ouverts et des services pour l'informatique interactive. (Voir Fig 4.3).



Figure 4.3 : Logo de l'IDE Jupyter.

3.2.1 Sklearn

Scikit-learn (sklearn) est une bibliothèque Python gratuite pour le ML. Il est développé par de nombreux contributeurs², notamment universitaires, par des instituts d'enseignement supérieur et de recherche français comme Inria³. Il fournit de nombreuses librairies d'algorithmes dans son Framework à implémenter, clé en main. Ces

bibliothèques sont particulièrement utiles pour les data scientists. Il comprend spécifiquement des fonctions d'estimation de forêts aléatoires, de régression logistique, d'algorithmes de classification et de machines à vecteurs de support. Il est conçu pour se coordonner avec d'autres bibliothèques Python gratuites, notamment NumPy et SciPy. (Voir Fig 4.4)



Figure 4.4 : Logo de bibliothèque Sklearn

3.3 Matrice de Confusion

Une matrice de confusion ou tableau de contingence sert à évaluer la qualité d'une classification. Elle est obtenue en comparant les données classées avec des données de référence qui doivent être différentes de celles ayant servi à réaliser la classification. Elle ne doit pas être confondue avec la table de performance qui sert à évaluer l'homogénéité thématique des zones-test servant à réaliser une classification assistée. (Voir Tableau 4.1)

		Valeur de prédiction	
		Positive	Négative
Valeurs réelles	Positive	VP	FN
	Négative	FP	VN

Tableau 4.1 : Exemple de Matrice de Confusion.

Table: Exemple de Matrice de Confusion.

Les paramètres d'évaluation sont définis comme suit :

- Vrai positive VP : classe positive classée comme étant positive.
- Vrai négative VN : classe négative classée comme étant négative.

- Faux positive FP : classe négative considérée positive.
- Faux négative FN : classe positive considérée négative.

Nous avons utilisé les mesures de performances suivantes pour évaluer et montrer l'efficacité de notre modèle :

La justesse (Accuracy) : La justesse ou (Accuracy en Anglais) est l'un des critères permettant d'évaluer les modèles de classification. Elle désigne la proportion des prédictions correctes effectuées par le modèle, l'équation représente la formule de calcul de cette mesure:

$$ACC = (VP + VN) / (VP + VN + FP + FN)$$

4-L'apprentissage du modèle proposé (Classification)

4.1-Présentation des ensembles d'apprentissage utilisés

Nous avons défini les variables indépendantes "Age, le sexe , Budget, Saison, et type de site" et les variables dépendantes "Destination", où nous avons préparé 3 ensembles d'apprentissage de données pour construire un modèle de prédiction selon les objectifs de prédiction tracés.

Ensemble d'apprentissage (BDD) : Cet ensemble se compose de 05 caractéristiques (Age, sexe, budget, saison et type de site) ainsi que la colonne du chapitre qui fait référence à destinations touristiques. Ce groupe est utilisé pour prédire le comportement touristique et choisir une destination touristique (Voir Fig 4.5), (Voir Fig4.6).

	Age	le sexe	Budget	Saison	type de site	Destination
0	45	femme	fort	Printemps	Historique	Qalat Bani Hammad
1	20	femme	faible	Printemps	Religieux	zawiya el hamel
2	17	femme	faible	Hiver	loisirs	Qalat Bani Hammad
3	42	homme	fort	Automne	loisirs	chalalt kamra
4	27	homme	Moyen	Automne	Religieux	zawiya el hamel
5	30	homme	Moyen	Printanière	Historique	Qalat Bani Hammad
6	32	femme	Moyen	Automne	Historique	Qalat Bani Hammad
7	31	homme	Moyen	Hiver	autre	Qalat Bani Hammad
8	40	femme	fort	Hiver	autre	zawiya el hamel
9	43	femme	fort	Hiver	Historique	Qalat Bani Hammad
10	48	homme	fort	Automne	loisirs	chalalt kamra

Figure 4.5 : Un groupe d'apprentissage pour prédire le comportement touristique (BDD1).



Figure 4.6 : Graphique de la variable dépendante.

4.2- L'importation et le partitionnement de la base de données

Nous identifions d'abord la base de données "Comportement touristique".

Variables indépendantes et dépendantes (Voir Fig 4.7) :

```
[3]: import pandas as pd
data = pd.read_csv("Dataset12.csv")
x = data_num_cat.drop('Destination', axis=1)
y = data_num_cat['Destination']
```

Figure 4.7: L'importation de l'ensemble d'apprentissage et la définition des variables.

Nous importons l'ensemble de données en utilisant le Framework Pandas. La trame de données est chargée dans la variable data. Ensuite nous définissons les variables indépendantes "Age, le sexe, Budget, Saison, et type de site et les variables dépendantes " Destination". Dans la base de données nous avons donné partie test 65% pour l'arbre de décision et 70% test pour Naïve Bayes par l'utilisation de la fonction «train_test_split» (Voir Fig 4.8).

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1, random_state=65)
```

Figure 4.8 : Code de Partitionnement de l'ensemble d'apprentissage

4.2.1. Arbre de décision (Decision tree)

Nous devons créer une instance de l'algorithme Naïve Bayes. Pour le créer, nous utilisons la classe DecisionTreeClassifier du package sklearn.tree ,pour créer une instance de l'algorithme. Après nous utilisons la méthode d'ajustement pour ajuster l'algorithme sur les données d'apprentissage. .(Voir Fig 4.9)

Après nous utilisons la classe confusion_matrix de sklearn. metrics pour calculer la matrice de confusion du modèle.(Voir Fig 4.10) .

```
[128]: from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion='entropy')
classifier.fit(x_train, y_train)
y_pred = classifier.predict(x_test)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print('Confusion Matrix:')
```

Figure 4.9 : Modèle Arbre de décision et matrice de confusion.

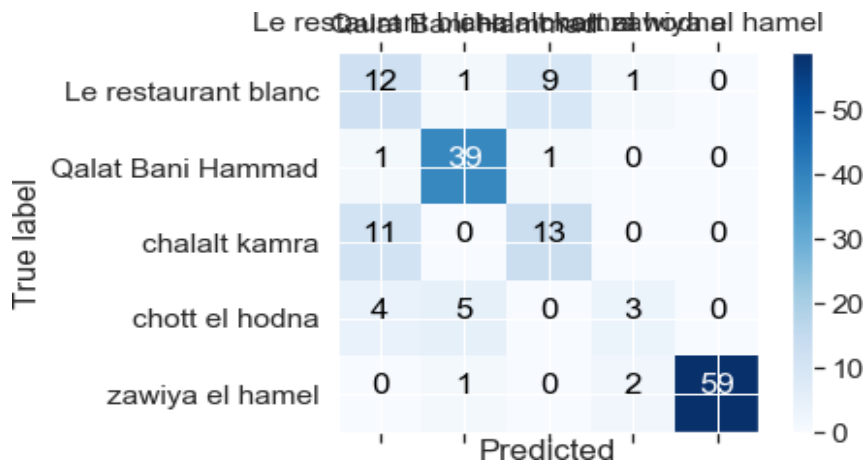


Figure 4.10 : Les résultats de matrice de confusion.

Par la matrice de confusion, les résultats indiqués dans le tableau ci-dessous ont été obtenus (voir tableau 4.2)

Attribue	BDD
Accuracy	80%

Tableau 4.2 : Les résultats obtenus en Arbre de Décision

Pour pouvoir obtenir une bonne représentation de notre résultat, nous utilisons cette méthode pour dessiner le graphe d'arbre (Voir Fig 4.11) , (Voir Fig 4,12) :

```
[157]: export_graphviz(classifier, out_file='tree01.dot', feature_names=x.columns,
                        class_names=['Qalat Bani Hammad', 'zawiya el hamel', 'chalalt kamra',
                                     'chott el hodna', 'Le restaurant blanc'], filled = True)
```

Figure 4.11 : un code pour dessiner l'Arbre de Décision.

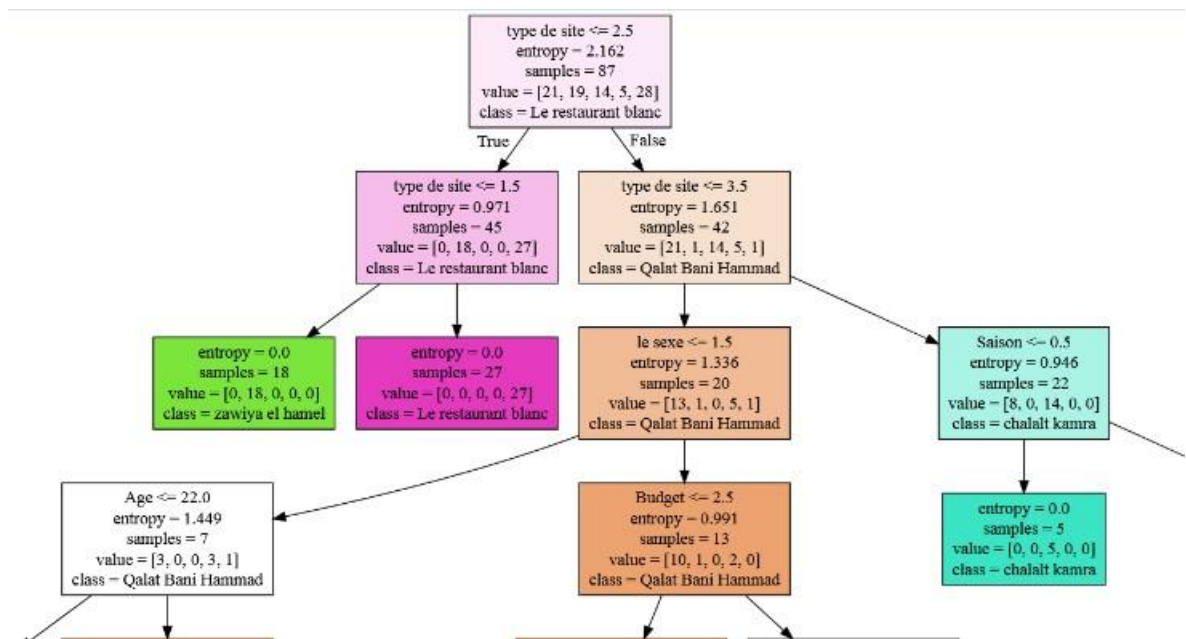


Figure 4.12 : Un extrait du graphe d'Arbre de Décision BDD

4.2.2. Naïf Bayésien :

Nous devons créer une instance de l'algorithme Naïve Bayes. Pour le créer, nous utilisons la classe GaussianNB du package sklearn. naive_bayes pour créer une instance de l'algorithme. Après nous utilisons la méthode d'ajustement pour ajuster l'algorithme sur les données d'apprentissage. (Voir Fig 4.13)

Après nous utilisons la classe confusion_matrix de sklearn. metrics pour calculer la matrice de confusion du modèle. (Voir Fig 4.14)

```
[86]: from sklearn.naive_bayes import GaussianNB
gaussian_nb = GaussianNB()
gaussian_nb.fit(x_train,y_train)
y_pred = gaussian_nb.predict(x_test)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print('Confusion Matrix:')
```

Figure 4.13 : Modèle Naïve Bayes et matrice de confusion.

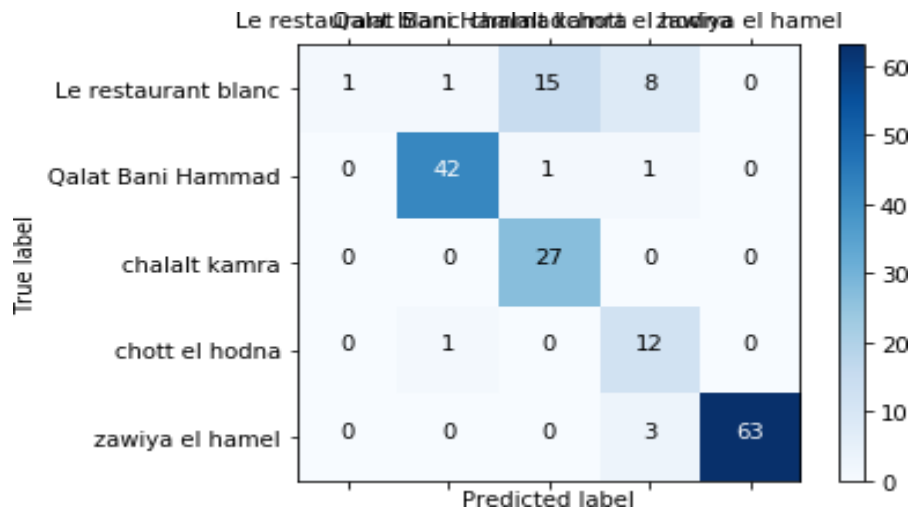


Figure 4.14 : Les résultats de matrice de confusion

Par la matrice de confusion, les résultats indiqués dans le tableau ci-dessous ont été obtenus (voir tableau 4.3).

Attribue	BDD
Accuracy	82%

Tableau 4.3 : Résultats obtenus par le modèle Naïve bayes.

4.2.3. Machines à Vecteurs de Support (SVM)

Nous devons créer une model de l'algorithme SVC sur les datasets . Pour le créer, nous utilisons :

- La noyau Kernel qui définit le type de fonction utilisé pour transformer l'ensemble de données de type 'rbf'.
- Le paramètre C qui définit la régularisation de l'erreur.
- Le coefficient du Kernel Gamma définit le degré d'ajustement du modèle aux

données d'entraînement, ce qui vous permet d'éviter le surajustement et le sousajustement.

- Degree est le degré de fonction polynomiale.

Ici, nous utilisons dans le modèle SVC le module GridSearchCV afin de tester un certain nombre de combinaisons de paramètres qui peuvent optimiser les performances de notre modèle. Pour le réglage des hyperparamètres, nous avons 2 paramètres à prendre en compte (paramètre C, paramètre gamma). (Voir Fig 4.15)

Pour obtenir la matrice de confusion nous avons utilisé la fonction «confusion_matrix» de la bibliothèque « sklearn » (Voir Fig 4.16)

```

from sklearn import svm
c = 1000
gamma = 0.01
kernel = 'rbf'
degree = 3
svm_model = svm.SVC(C=c, gamma=gamma, kernel=kernel, degree=degree)
svm_model.fit(x_train, y_train)
y_pred = svm_model.predict(x_test)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print('Confusion Matrix:')
print(cm)
    
```

Figure 4.15: Modèle SVM et matrice de confusion.

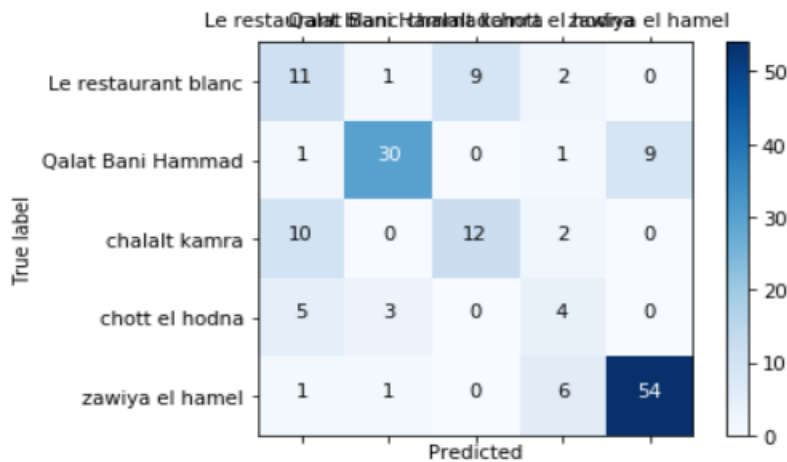


Figure 4.16 : Les résultats de matrice de confusion

Par la matrice de confusion, les résultats indiqués dans le tableau ci-dessous ont été obtenus (voir tableau 4,4).

Attribue	BDD
Accuracy	68%

Tableau 4.4: Résultats obtenus par SVM

4.3. Résultats d'évaluation des approches d'apprentissage automatique sur les donnée :

Commençons, tout d'abord par comparer notre méthode de classification aux différentes approches existantes, en termes de Accuracy pour chaque Ensemble d'apprentissage, nous avons déroulé les algorithmes suivants : Arbre de décision et SVM la taille du test est égale à 0.65 , Pour le Naïf bayes ,la taille de test égala à 0.7.

Ensuite, Nous avons calculé l'Accuracy des approches par rapport à notre datasets comme suivants (voir tableau 4.5):

	Arbre de décision	SVM	Naive bayes
Accuracy	80%	68%	82%

Tableau 4.5: Pourcentage de Accuracy pour BDD.

Après avoir confronté ensemble d'apprentissage BDD aux algorithmes, nous avons constaté que Naive Bayes ont donné un bon résultat par rapport à l'Arbre dedécision et svm avec Accuracy 82%.

Conclusion :

D'après les résultats de l'approche proposée, il a été constaté que les résultats de précision sont moyens, et cela n'est pas suffisant pour obtenir un modèle idéal pour prédire les comportement touristique. Pour obtenir de meilleurs résultats, il est suggéré d'augmenter le nombre d'années qu'il faut pour la formation et les tests.

À l'avenir, nous proposons de tester l'approche proposée en utilisant Apprentissage en ligne dans le cadre du système de gestion du tourisme et du comportement touristique.

Conclusion générale

Dans ce mémoire, dans le premier chapitre. Nous avons appris sur le tourisme public et le tourisme en Algérie, de ses genres, l'objectif du tourisme et son importance sur l'économie du pays. Le tourisme est devenu le principal soutien de l'économie, de la prospérité et du développement du pays. Nous avons aussi appris à connaître le tourisme et le comportement des touristes.

Dans le deuxième chapitre, le rôle de l'informatique, en particulier de l'intelligence artificielle, dans la prévision des comportements touristiques et dans le choix des touristes pour une destination touristique particulière, afin d'aider la Direction du tourisme et les agences touristiques à comprendre les comportements touristiques et à répondre aux besoins en services pour attirer les touristes. Pour cela, nous avons détaillé les concepts du domaine de l'intelligence artificielle et de la DM.

Dans le troisième chapitre, nous lisons une partie de l'étude et des méthodes liées au comportement des touristes pour comprendre les opinions des chercheurs sur le sujet.

Notre approche et l'ingénierie proposée sont détaillées au quatrième chapitre, basé sur l'utilisation de trois types d'algorithmes précédemment formés : Bayes naïve et Arbre de décision, notre sélection de ces types d'algorithmes sur les algorithmes Plus précis et largement utilisé que d'autres algorithmes d'apprentissage supervisé. Nous avons prouvé l'efficacité de nos modèles pour prédire les comportements touristiques tels que Naïve Bayes était le modèle le plus efficace dans les trois ensembles de données avec une prévision dépassant 82%.

Nous souhaitons dans l'avenir, que les institutions touristiques en Algérie, et plus particulièrement à Msila, donnent plus d'intérêt à la collecte des données relatives au touriste, ceci pour réaliser des modèles prédictifs plus efficaces. Les études futures peuvent également utiliser d'autres modèles et les tester pour choisir le meilleur qui donne la grande accourcie.

Bibliographie

- [1] Liu, S., & Fan, J. (2013). Tourist Behavior Pattern Mining Model Based on Context. *Hindi Publishing Corporation*, 13.
- [2] Abdellilah, N. (n.d.). CLASSIC AND DATAMINING MODELS MOST USED IN EVALUATION AND PREDICTION OF THE PERFORMANCE OF .
- [3] Bousbia, L. M., & Tedjani, M. (2022). *Machine learning based system*. EL OUED : UNIVERSITY OF ECHAHID HAMA LAKHDER EL OUED .
- [4] Dale, F., & Brian, M. (1999). A model of tourist information search behavior. *journal of travel research*.
- [5] Gallo, G., Signorello, G., & Farinella, G. M. (2017). *Exploiting Social Images to Understand Tourist*. Ital: University of Catania, Catania, Italy.
- [6] Hakeem , B., & Maryam , B. (2020). Monitoring the direction of the tourists of Ouarglacity to their destination based on reference groups. *magazine entreprise* , 321.
- [7] HAMMICHE, M. (2020). *Les enjeux de marketing des destinations touristique* . BEJAIA: UNIVERSITE ABDERAHMANE MIRA DE BEJAIA.
- [8] Kalm, S. (2022). The role of tourist awareness in preserving environmental and . *Cahiers de recherche scientifique, volume 10, numéro 1.*, 356.
- [9] Núria, G., Thomas, F., Delphine, B., & Cédric, F. (2014). *Un modèle pour prédire la complexité lexicale et graduer les mots*. Marseille: Université Catholique de Louvain, Aix Marseille Université.
- [10] Shah , J., & Miah , H. Q. (2016). A big data analytics method for tourist behaviour. *Accepted Manuscript*, 378.
- [11] Suthaharan, S. (2016). Machine learning models and algorithms for big data . *Integr. Ser. Inf. Syst*, 36.
- [12] Thinhinane, A., & Belaid, A. (2017). *Etude comparative des choix des destinations touristiques: cas des complexes touristiques Algérie – Tunisie*. TIZI OUZOU: UNIVERSITE MOULOUD MAMMERI DE TIZI OUZOU.
- [13] Xu, S., Huang, Q., & Zou, Z. (2023). Spatio-Temporal Transformer Recommender. *ISPRS Int.*

- [14] جامعة محمد الصادق بن يحي جيجل : جيجل .العوامل المحددة اختيار السائح الوجهات السياحية . (2017). و , بوكروح & , ن. اودينة
- [15] ش بوجمة ، و م بلمقدم . (2020). تقييم النشاط السياحي في الجزائر و اثره على النمو الاقتصادي . البشائر الاقتصادية ، 508.
- [16] جامعة شهيد حمه لخضر الوادي : الوادي .سلوك المستهلك السياحي الجزائري . (2020). ع , سلطان & , ح , مراد , م , نبيل
- [17] Zighed. D.A and Rakotomalala. R. Data Mining,2002
- [18] Lemberger .P, Batty. M, Morel. M and Raffaëlli. JL. Big Data et Machine Learning,2015
- [19] Panagiotis Korvesis, le 21 novembre 2017, « Apprentissage Automatique pour la Maintenance Predictive dans le Domaine de l'Aviation », Thèse de doctorat, sous la direction de Michalis Vazirgiannis, Paris, de l'Université Paris-Saclay préparée à l'école Polytechnique, 139p.
- [20] Roger Bagra, Valentine Fontana, Wee-hyong Tok, 2014, « Predictive analytics Microsoft azure machine learning», edition 1,188p.
- [21] Innocent Mateyaunga, 2020, « Prédictive Maintenance Using Machine Learning », la thèse de master, sous la direction de Hadj Abdelkader, Faculté de technologie de l'université de Tlemcen, 61p
- [22] Cortes, C., & Vapnik, V. 1995, "Support-vector networks", Machine learning,20(3), 273-297.
- [23] Bassem Besbes. Intégration de methodes de representation et de classification pour la détection et la reconnaissance d'obstacles dans des scenes routieres. Computer Science. INSA de Rouen,2011. French. <NNT: 2011ISAM0007>. <tel 00633109>
- [24] python.doctor. (2019). Apprendre le langage de programmation python. Retrieved 04 29, 2020, from python.doctor: <https://python.doctor/>
- [25] G. Gibran, "Introducción al machine learning en Python." <https://naps.com.mx/blog/introduccion-al-machine-learning-en-python/>. Accessed: 2022-03-20. 7
- [26] G. Ligdi, "Clasificación de Machine Learning." <https://aprendeia.com/clasificacion-de-machine-learning/>. Accessed: 2022-02-20. 12
- [27] Direction du tourisme et de l'artisanat M'sila, <https://msila.mta.gov.dz/fr/>

Résumé

Le modèle d'analyse du comportement touristique est conçu pour comprendre et interpréter le comportement touristique afin de fournir des informations précieuses aux professionnels du tourisme. Ce modèle utilise des techniques avancées d'apprentissage automatique pour analyser les données touristiques provenant de diverses sources, comme les médias sociaux, les sites de voyage, les réservations d'hôtel, les commentaires des utilisateurs, etc.

En analysant les données, le modèle peut faire des recommandations aux touristes, en proposant des destinations, des hébergements, des activités et des expériences qui correspondent à leurs préférences individuelles. Il peut également aider les professionnels du tourisme à mieux cibler leurs efforts de marketing, à identifier les segments de marché prometteurs et à concevoir leurs offres en conséquence.

L'apprentissage supervisé est l'une des branches les plus importantes et exploitées de l'apprentissage automatisé, dépendant de l'apprentissage et de l'acquisition supervisés Expériences. Il imite la façon dont les gens pensent et apprennent. Il est considéré comme une machine de soutien Bayes naïve et arbre Décision est l'un des premiers algorithmes d'apprentissage supervisé. Ils ont donné des résultats très efficaces, notamment en matière de classification et de régression.

Dans ce travail, nous avons proposé une nouvelle approche basée sur une partie de l'intelligence artificielle qui est l'éducation supervisée pour prédire le comportement touristique et cette approche repose sur deux algorithmes : classification naïve Bayes et arbre de décision.

Mots clés : tourisme, comportement touristique, la prédiction.

Abstract

The tourism behaviour analysis model is designed to understand and interpret tourism behaviour in order to provide valuable information to tourism professionals. This model uses advanced machine learning and natural language processing techniques to analyze tourism data from various sources, such as social media, travel sites, hotel reservations, user feedback, etc.

By analyzing the data, the model can make recommendations to tourists, offering destinations, accommodations, activities and experiences that match their individual preferences. It can also help tourism professionals better target their marketing efforts, identify promising market segments and design their offerings accordingly.

Supervised learning is one of the most important and exploited branches of automated learning, dependent on supervised learning and acquisition experiences. It imitates the way people think and learn. It is considered a support machine Naive Bayes and Decision Tree is one of the first supervised learning algorithms. They have been very effective, particularly in terms of classification and regression.

In this work, we have proposed a new approach based on a part of artificial intelligence which is supervised education to predict tourist behavior and this approach is based on two algorithms: naive Bayes classification and decision tree.

Keywords: tourism , tourist behavior, prediction.

ملخص:

تم تصميم نموذج تحليل السلوك السياحي لفهم وتفسير السلوك السياحي لتوفير معلومات قيمة للمهنيين السياحيين. يستخدم هذا النموذج التعلم الآلي المتقدم وتقنيات معالجة اللغة الطبيعية لتحليل بيانات

السياحة من مصادر مختلفة، مثل وسائل التواصل الاجتماعي ومواقع السفر وحجوزات الفنادق وتعليقات المستخدمين وما إلى ذلك.

من خلال تحليل البيانات، يمكن للنموذج تقديم توصيات للسياح، وتقديم الوجهات وأماكن الإقامة والأنشطة والتجارب التي تتناسب مع تفضيلاتهم الفردية. يمكن أن يساعد أيضًا المتخصصين في السياحة على استهداف جهودهم التسويقية بشكل أفضل، وتحديد قطاعات السوق الواعدة وتصميم عروضهم وفقًا لذلك.

التعلم الخاضع للإشراف هو أحد أهم فروع التعلم الآلي وأكثرها استغلالاً، ويعتمد على التعلم والاكتماب الخاضعين للإشراف تجارب. إنه يقلد الطريقة التي يفكر بها الناس ويتعلمون. تعتبر آلة دعم تصنيف بايزي الساذج و شجرة القرار هي واحدة من أولى خوارزميات التعلم الخاضعة للإشراف. لقد كانت فعالة للغاية، لا سيما من حيث التصنيف والانحدار. في هذا العمل، اقترحنا نهجًا جديدًا يعتمد على جزء من الذكاء الاصطناعي يخضع للإشراف التعليمي للتنبؤ بالسلوك السياحي ويستند هذا النهج إلى ما قبل الخوارزمية المدرب: تصنيف بايز الساذج وشجرة القرار.

الكلمات المفتاحية : السياحة , السلوك السياحي , التنبؤ .