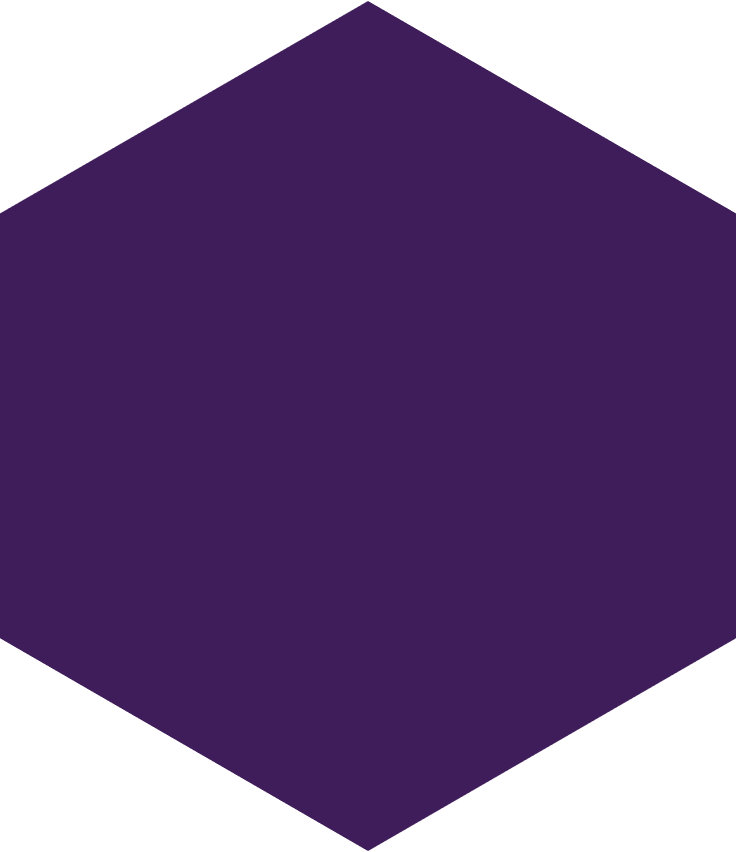
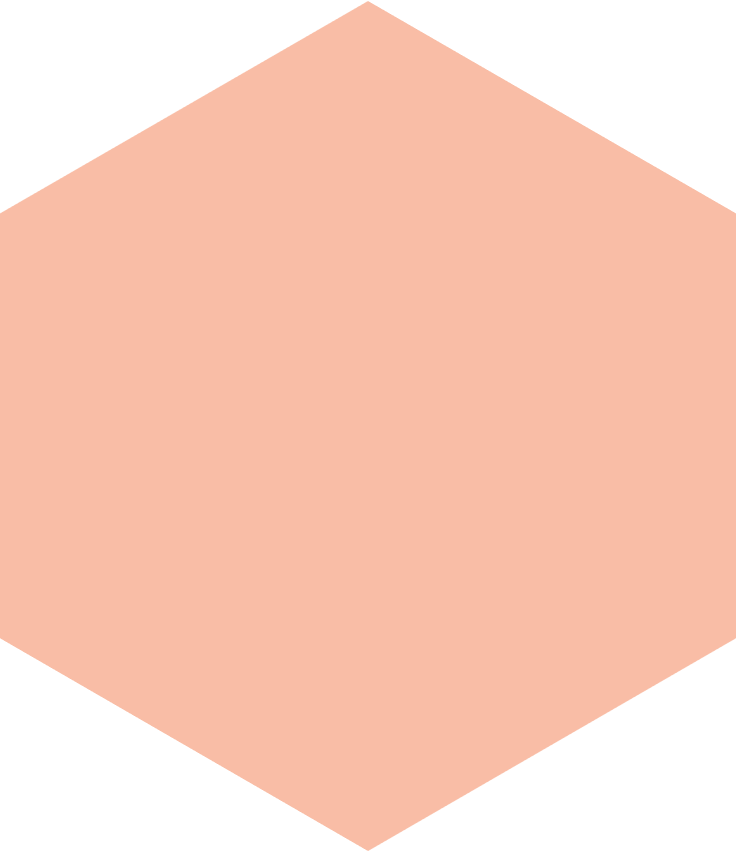


|  |
| --- |
| **Prévision de nombre des étudiants présents dans les restaurants universitaires** |
| Construction de jeu de données |
| Le présent rapport permet de décrire la construction de jeu de données, en détaillant les différents attributs et variables influençant sur l’entrainement du modèle de prévision. Une démarche de génération pseudo-aléatoire de données a été abordée également. |
|  |





|  |  |  |  |
| --- | --- | --- | --- |
| Introduction |  | Objectifs • • • |  |
| - Etudier l’environnement des restaurants universitaires et les différents facteurs externes.- Etudier la population concernée (les étudiants) Et c'est en étudiant la situation de vie de l'étudiant et les facteurs les plus importants affectant l'intention de l'étudiant au restaurant universitaire. - Etudier les évènements, les jours fériés et les alertes épidémiques permettant de faire une différence dans le nombre des étudiants présents.  - Etudier les données météorologiques.  - Construire la table finale de la prévision. |  |
|  |  |
| Définition du problème : La deuxième partie de notre projet de fin d’étude consiste à développer des modèle intelligent afin de prévoir le nombre des étudiants présents dans les restaurants universitaires. Pour cela un ensemble d’algorithmes ont été étudiés afin de sélectionner le modèle le plus adéquats. Les modèles ont été partagés en deux types : modèles stochastiques et modèles des réseaux de neurones artificiels.  Cependant, une étape très importante doit précéder cette étape de modélisation, il s’agit bien de la construction de jeu de données, afin d’identifier tous les attributs qui influencent sur l’apprentissage du modèle et sur le score final de ce dernier. Démarche à suivre : Dans ce qui suit, nous allons étudier attentivement les attributs construisant la table finale de prévision, pour cela, nous avons tracer un certain nombre de tables intermédiaires qui sont :   * **La table Menu.** * **La table Restaurant.** * **La table Evènements.** * **La table Population.** * **La table météo.**   Pour chaque table, nous détaillerons les attributs associés, ainsi que la manière de les représenter.  La table finale de prévision sera construite en faisant la jointure entre les différentes tables citées précédemment. |  |

|  |  |
| --- | --- |
| Définition des tables |  |
|
|  |
|  |

## La table ‘’Menu’’ :

La table ‘’Menu’’ est une table qui contient les informations décrivant le repas en mentionnant l’identifiant du restaurant, la date, le type du repas (midi ou soir) et finalement le plat consommé.

Il est à noter que certains restaurants ne proposent pas des repas soir comme c’est le cas de la majorité des restaurants des université ou des écoles supérieur.

Nous voyons que la table ‘’Menu’’ contribue dans la construction de notre jeu de données, ceci peut être justifié, le fait que certains étudiants préfèrent un plat par rapport à un autre ce qui va influencer d’une manière directe sur le nombre total des présents dans le restaurant.

Voici la structure de la table ‘’Menu’’ avec des exemples d’enregistrements :

|  |  |  |  |
| --- | --- | --- | --- |
| IdRestaurant | Date | Type | Plat |
| ESI016 | 21/04/2022 | Midi | Riz |
| Bouraoui016 | 21/04/2022 | Soir | Soupe de poisson |

Définition des attributs :

* **IdRestaurant :** une chaine de caractère qui représente l’identifiant du restaurant, elle doit être unique.
* **Date**: de type date, c’est la date du jour qui correspond au repas.
* **Type :** C’est le type du repas (midi ou soir). Dans ce cas nous pouvons opter pour une modélisation binaire de telle sorte que l’attribut prendra la valeur 1 s’il s’agit d’un repas de midi ou 0 pour un repas de soir.
* **Plat :** Une chaine de caractère qui décrit le plat de la journée.

***Remarque : l’alimentation de cette table se fait via l’interface web par les employés des restaurants universitaires***.

## Démarche à suivre :

## La table ‘’Restaurant’’ :

La table ‘’Restaurant’’ est une table qui décrit le restaurant ainsi que son environnement, en mentionnant toute sorte d’information qui pourra influencer sur le nombre d’étudiants présents dans les restaurant universitaires.

Cette table va contenir tous les restaurants des établissements (cités universitaires, universités, écoles supérieures).

Le fait de mettre tous les restaurants dans une seule table pourra rendre l’algorithme de prévision plus performant en essayant de trouver des restaurants qui ont les mêmes caractéristiques.

Voici à quoi ressemble la table ‘’Restaurant’’ :

|  |  |  |  |
| --- | --- | --- | --- |
| IdRestaurant | Effectif | Nb\_src\_Alim | Prix\_moy\_Rep |
| ESI016 | 450 | 5 | 210 |
| Bouraoui016 | 800 | 3 | 190 |

Définition des attributs :

* **IdRestaurant :** une chaine de caractère qui représente l’identifiant du restaurant, elle doit être unique.
* **Effectif**: représente la capacité maximale du restaurant.
* **Nb\_src\_Alim :** C’est le nombre de sources d’alimentation qui se trouvent à coté du restaurant, il peut s’agir de fast-food, des magasins, des restaurants externes …
* **Prix\_moy\_Rep :** C’est le prix moyen d’un repas dans les sources d’alimentation externes.

***Remarque : l’alimentation de cette table ne pourra pas être fait via l’interface web, par conséquent, il nous faut un effort supplémentaire afin de saisir ce type d’informations.***

## Démarche à suivre :

## La table ‘’Evènement ’’ :

La table ‘’Evènement’’ est la table qui contient le planning des vacances, des jours fériés et des fêtes religieuses ainsi que les alertes épidémiques possible.

Un évènement peut influencer sur la variable cible selon sont type.

Voici à quoi ressemble la table ‘’Evènement’’ :

|  |  |
| --- | --- |
| Date | Type |
| 20/04/2022 | Ramadan |
| 02/05/2022 | Jour férié |

Définition des attributs :

* **Date :** C’est la date de l’événement.
* **Type :** C’est le type de l’évènement, il peut s’agir d’un jour férié, de vacances ou de mois de ramadan. Un codage est spécifié à chaque type d’événement.

Nous pouvons opter pour le codage suivant : 0 : jour férié, 1 : vacances, 2 : Ramadan, 3 : alerte épidémique.

***Remarque : l’alimentation de cette table se fait via l’interface web du tableau de bord par les décideurs du ministère de l’enseignement supérieur et de la recherche scientifique lors de la planification des jours fériés et des congés universitaires.***

## La table ‘’Population’’ :

La table ‘’Population’’ décrit précisément le niveau de vie des bénéficiaires du service de la restauration qui sont les étudiants.

Cette table permet de résumer les différents paramètres influençant sur la variable cible en fonction de la situation personnelle de l’étudiant ou même de sa famille.

Voici à quoi ressemble la table population :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IdRestaurant | IdEtudiant | Date | Heberg | Bourse | Rev\_parents |
| ESI016 | 21/0232 | 20/04/2022 | Interne | OUI | 65 000 |
| Bouraoui016 | 21/0242 | 20/04/2022 | Externe | NON | 48 000 |

Définition des attributs :

* **IdRestaurant :** une chaine de caractère qui représente l’identifiant du restaurant, elle doit être unique.
* **IdEtudiant :** une chaine de caractère qui représente l’identifiant de l’étudiant, elle doit être unique.
* **Date :** C’est la date du jour ou l’étudiant à pris son repas
* **Heberg :** Un attribut pour spécifier si un étudiant est interne ou externe. (Il est toujours possible d’opter pour un codage binaire à titre d’exemple : 0 pour interne et 1 : externe).
* **Bourse :**Un attribut pour spécifier si un étudiant bénéfice de la bourse ou pas. (Il est toujours possible d’opter pour un codage binaire à titre d’exemple : 0 pour dire qu’il ne bénéfice pas et 1 : si oui).
* ***Rev\_parents :*** C’est le revenu moyen des parents de l’étudiant. Il peut d’agir des deux parents de l’étudiant ou un d’eux dans le cas ou il n’est pas salarié\* ;

***Remarque : l’alimentation de cette table pourra se faire via l’interface web de l’application web à travers les coordonnées de l’étudiant lors de son inscription.***

## Démarche à suivre :

## La table ‘’Météo’’ :

La table ‘’Météo’’ contient les données météorologiques.

Cette table peut influencer sur le nombre d’étudiants présents dans les restaurants.

Voici à quoi ressemble la table ‘’Météo’’ :

|  |  |
| --- | --- |
| Date | Météo |
| 21/04/2022 | Ensoleillé |
| 22/04/2022 | Pluvieux |

Définition des attributs :

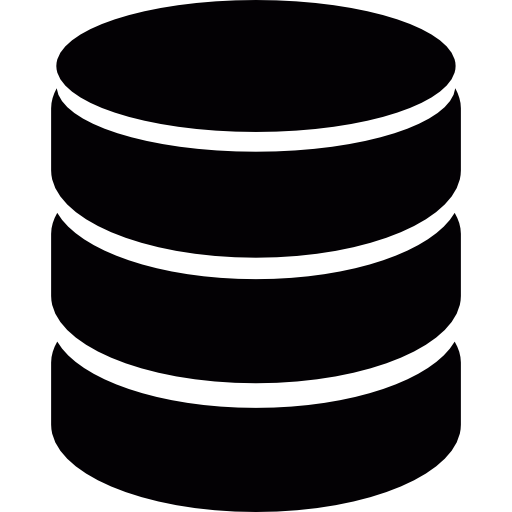
* **Date:** date de la journée.
* **Météo :** des données météorologiques codés comme suit : (0 : pluie, 1 : Ensoleillé…)

***Remarque : l’alimentation de cette table pourra se faire à travers les apis de Google weather.***

La table de prévision, c’est la table sur laquelle nous allons entrainer nos différents modèles d’apprentissage, elle va être constitué en mettant une jointure entre les différentes tables décrites précédemment ; cette jointure peut se faire à travers l’identifiant du restaurant.

Une étape de sélection d’attributs ensuite sera faite afin de choisir les variables ou les attributs les plus pertinents qui influencent sur le score de l’algorithme d’apprentissage.

La table de prévision :



Dans cette partie, nous allons expliquer la manière dont nous avons généré les tables décrites précédemment.

Il est important de suivre une logique afin que notre jeu de données soit cohérent.

Nous avons généré un jeu de donnée d’une période de 3ans.

## Caractéristique générale du jeu de données :

Notre jeu de données doit comporter un historique au minimum de 3ans, de **01 septembre 2019 au 01 septembre 2022.**

Le jeu de données doit être construit à la base des tables définies dans le point précédent, en joignant les différentes tables selon l’identifiant de chacune et sélectionnant les attributs les plus important.

Ensuite, une fois le jeu de données est prêt, nous allons le diviser en deux parties, une partie qui va être consacrer pour l’entrainement de notre modèle de prévision, et une autre pour effectuer le test et calculer le score du modèle.

Il est clair que la partie de l’apprentissage doit être plus volumineuse que celle du test.

Voici comment il est effectué le partage du jeu de données :

## Démarche à suivre :

Génération pseudo aléatoire des tables

## Génération de la table météo :

Rappelons la structure de la table météo :

|  |  |
| --- | --- |
| Date | Météo |
| 21/04/2022 | Ensoleillé |
| 22/04/2022 | Nuageux |

Nous avons choisi 3 valeurs pour l’attribut ‘’meteo’’ qui sont : **Ensoleillé, Nuageux et Mauvais.**

Pour chaque mois nous avons affecté une probabilité à chacune de ces valeurs, par exemple le mois de septembre, 75% des cas il est ensoleillé, 25% nuageux et 0% mauvais.

Voici le code python associé :

Type = ['Ensoleillé' , 'Nuageux' , 'Mauvais']

weightsSep = [0.75,0.25,0]

weightsOct = [0.25,0.75,0]

weightsNov = [0.2,0.8,0.2]

weightshiv = [0,0.3,0.7]

weightsmars = [0.1,0.2,0.7]

weightsAvr = [0.25,0.4,0.35]

weightsmai = [0.5,0.4,0.1]

weightsete = [0.95,0.05,0]

Et voici un aperçu sur la table météo :

## Démarche à suivre :

## Génération de la table évènement :

Rappelons la structure de la table évènement :

|  |  |
| --- | --- |
| Date | Type |
| 20/04/2022 | Ramadan |
| 02/05/2022 | Jour férié |

Nous avons choisi 3 valeurs pour l’attribut ‘’Type’’ qui sont :

* **Normal :** Pour signifier qu’il n’y a aucun évènement particulier dans ce jour.
* **Jour férié :** pour signifier qu’il s’agit d’un jour férié**.**
* **Ramadan :** pour signifier qu’il s’agit d’un jour du mois de ramadan**.**

***Remarque : Ces informations ont été remplies manuellement d’une manière correcte.***

Et voici un aperçu sur la table Evènement :

## Démarche à suivre :

## Génération de la table évènement :

Rappelons la structure de la table évènement :

|  |  |
| --- | --- |
| Date | Type |
| 20/04/2022 | Ramadan |
| 02/05/2022 | Jour férié |

Nous avons choisi 3 valeurs pour l’attribut ‘’Type’’ qui sont :

* **Normal :** Pour signifier qu’il n’y a aucun évènement particulier dans ce jour.
* **Jour férié :** pour signifier qu’il s’agit d’un jour férié**.**
* **Ramadan :** pour signifier qu’il s’agit d’un jour du mois de ramadan**.**

***Remarque : Ces informations ont été remplies manuellement d’une manière correcte.***

Et voici un aperçu sur la table Evènement :

## Démarche à suivre :

## Génération de la table Menu:

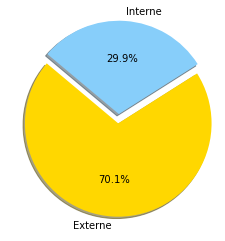
Rappelons la structure de la table évènement :

|  |  |  |  |
| --- | --- | --- | --- |
| IdRestaurant | Date | Type | Plat |
| ESI016 | 21/04/2022 | Midi | Riz |
| Bouraoui016 | 21/04/2022 | Soir | Soupe de poisson |

Nous avons rempli l’attribut Plat d’une manière périodique avec les plats suivant : **Lentilles – Loubia – Riz – Spaghetti – Couscous.**

Et voici un aperçu sur la table Evènement :

## Démarche à suivre :



## Génération de la table Population :

Rappelons la structure de la table Population:

|  |  |  |  |
| --- | --- | --- | --- |
| IdEtudiant | Heberg | Bourse | Rev\_parents |
| 21/0232 | Interne | OUI | 65 000 |
| 21/0242 | Externe | NON | 48 000 |

Et voici un aperçu sur la table ‘’population’’:



Nous avons généré un ensemble de 1000 étudiants suivant les règles suivantes :

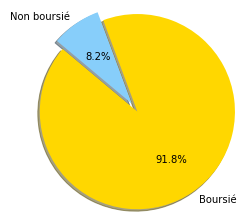
* ***L’attribut Hébergement :***

Sur 1000 étudiants, nous avons classé 70% de la

population comme des étudiants externes

et 30% comme des étudiants internes.

## Démarche à suivre :

* ***L’attribut Bourse :***

L’attribut Bourse a été généré suivant le revenu moyen des parents, de telle sort que si le revenu est supérieur à 150000, l’étudiant ne pourra pas bénéficier de la bourse.

* ***L’attribut Rev\_par :***

L’attribut ‘’Rev\_par’’ qui veut dire le revenu moyen des parents de l’étudiant généré de la manière suivante :

Nous avons classifié les salaires par catégorie, et pour chaque catégorie nous donné une probabilité d’apparition comme suite :

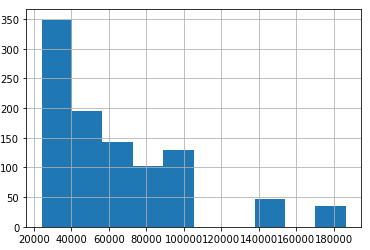
salaire = [random.randrange(18000,25000), random.randrange(25000,40000) , random.randrange(40000,50000), random.randrange(50000,60000),

random.randrange(60000,80000),random.randrange(80000,100000), random.randrange(100000,120000), random.randrange(120000,150000), random.randrange(150000,200000)]

poids = [0.20,0.15,0.2,0.15,0.1,0.06,0.06,0.05,0.03]

Nous avons donné par exemple un poids plus important aux revenus qui varient entre 18000 et 25000 vu que le salaire de base en Algérie et proche de 20000 DA.

Voici en représentation graphique la distribution de salaires des parents de notre population :



## Notre jeu de données, bientôt arrivé !

Maintenant, il va falloir faire une jointure entre les différentes tables afin de construire la dernière variable qui est la variable de sortie, cette variable va dépendre impérativement des différents attributs de la table.

Voici d’abord un aperçu de la table après une jointure, sachant que nous construisant une table sur une période de 3ans et en choisissant un seul restaurant doté de 1000 étudiants nous obtiendrons une table avec 1080000 lignes.

