

# MÉMOIRE DE PROJET DE FIN D'ANNÉE

Présenté à

L'ÉCOLE NATIONALE DES SCIENCES APPLIQUÉES AL HOCEIMA

Spécialité : Génie Informatique

Titre du PFA :

Recherche et développement d'une plateforme web de prédiction de la réussite au sein de l'ENSA-Hoceima

Réalisé par :

- Abdessamad FELOUACH
- Abdessalam BADOUI

Encadré par :

Pr. Salmani Abdelhafid

Déposé le : 30 juillet 2020

Devant le jury composé de :

Pr. Abdelhafid Salmani

Pr. El Wardani Dadi

ENSA Al Hoceima

ENSA Al Hoceima

*Année Universitaire : 2019/2020*

“Data is a precious thing and will last longer than the systems themselves.”

Tim Berners-Lee,

Inventeur du World Wide Web.

A nos chers parents

A nos amis et nos collègues

A tous nos enseignants avec notre considération

Et à tous ceux qui nous ont assistés, dans la réalisation et le bon déroulement de ce travail.

# Remerciements

On remercie infiniment nos professeurs au sein de l'Ecole Nationale des Sciences appliquées d'Al Hoceima, notre encadrant **M.Salmani Abdelhafid**, pour son suivi, sa disponibilité ainsi que ses nombreux conseils bienveillants, forts et utiles pendant toute la période de la réalisation du projet. Nos vifs remerciements s'adressent également aux membres du jury qui ont accepté de juger notre travail. Aussi, on tient à remercier tous les étudiants au sein de notre école pour leur collaboration, toute personne ayant contribué de près ou de loin à l'aboutissement de notre travail.

# Résumé

Ce rapport est le résultat d'un travail collectif que nous avons effectué dans le cadre du projet de fin d'année de la filière Génie Informatique au sein de l'Ecole Nationale Des Sciences Appliquées d'Al Hoceima (ENSAH), dont le sujet est de rechercher et analyser les données des élèves ingénieurs afin de réaliser une plateforme de prédiction et de recommandation pour le but d'améliorer le parcours scolaires des étudiants de l'école national des sciences appliquées d'Al-Hoceima

La Science des données (data science) : est une science interdisciplinaire s'appuyant sur des méthodes scientifiques, des algorithmes, des processus et autres systèmes afin d'exploiter de grands ensembles de données. Les data scientists associent ainsi plusieurs compétences, notamment des connaissances en informatique, statistiques et commerce pour analyser des données collectées auprès des clients ou bien d'autres sources à l'aide de capteurs, de leurs Smartphones, de leurs habitudes de navigation sur internet, etc.

# Abstract

This paper about building a recommendation and prediction web platform focused on improving student performance prediction , based on their personal and academic performance characteristics, prediction students 'performance will alert the learner to know about their performance and gives as a chance to improve their performance in the future , the dataset used for the research purposes includes data about ENSAH Students during the current semester , Data Mining algorithms is used to predict the student performance .

# Table des matières

Remerciements .....	iv
Résumé .....	v
Abstract.....	vi
Table des matières .....	vii
Liste des figures .....	x
Liste des tableaux .....	xi
Introduction générale.....	1
Contexte Du Projet .....	3
1.1    Présentation de Projet .....	3
1.1.1    Présentation de l'école nationale des sciences appliquées d'Al Hoceima (ENSAH) .....	3
1.1.2    Notions de base .....	4
1.1.2.1    Data Science .....	4
1.1.2.2    La science de données au service de l'université .....	5
1.1.3    Problématique .....	5
1.1.4    Contexte du Projet.....	6
1.2    Conduite et gestion de Projet.....	7
1.2.1    Contraintes de Projet .....	7
Contraintes techniques.....	7
1.2.2    Gestion de Projet avec Trello.....	8
1.2.3    Gestion des versions avec Git .....	9
1.3    Conclusion.....	11
Le Cycle De Vie Du Projet .....	12
2.1    La Collecte De Données.....	13
2.1.1    Représentation d'échantillon .....	14
2.2    Compréhension Des Données.....	14
2.2.1    Les Types De Données .....	15
2.2.1.1    Les Variables Qualitatives .....	15

2.2.1.2	Les Variables Quantitatives .....	16
2.3	Préparer Les Données .....	16
2.3.1	Nettoyage et transformation Des Données .....	18
2.4	Le « Machine Learning » : Connaissances nécessaires .....	19
2.4.1	Objectif et caractéristiques du problème d'apprentissage .....	20
2.4.2	Types d'apprentissage .....	21
2.4.2.1	Apprentissage supervisé.....	21
2.4.2.2	Apprentissage non supervisé :.....	22
2.5	Conclusion.....	24
Étude Fonctionnelle.....		25
3.1	Analyse Fonctionnelle.....	25
3.1.1	Définir les besoins.....	26
3.1.1.1	La Bête à Cornes.....	26
3.1.1.2	Diagramme Pieuvre .....	27
3.1.2	Identifications des différents modules.....	28
3.1.3	Diagrammes de Séquence.....	28
3.2	Conclusion.....	30
Etude Technique.....		31
4.1	Architecture technique du projet .....	32
4.2	Choix logiciels.....	33
4.2.1	Langage Python.....	33
4.2.2	Framework Flask.....	34
4.2.3	Numpy .....	34
4.2.4	Scikit-learn.....	34
4.2.5	Pandas.....	35
4.2.6	Matplotlib .....	35
4.2.7	Pickle .....	35
4.3	Conclusion.....	36
5	Réalisation Et Mise En Œuvre.....	37
5.1	Outils de développement.....	37
5.1.1	Training Application Python Back App.....	37



5.1.1.1	IDE Python.....	38
5.1.1.2	Python Distribution : Anaconda.....	39
5.1.1.3	Base de données : Microsoft Excel.....	39
5.1.1.4	Comma-separated values : CSV.....	40
5.1.1.5	Power BI.....	41
5.1.2	ENSAH Prédiction Front.....	43
5.1.2.1	Sublime Text.....	43
5.2	Présentation de la plateforme ENSAH Prédiction.....	44
5.3	Conclusion.....	45
Conclusion générale .....		53
1	Bibliographie.....	54

# Liste des figures

Figure 1 l'école nationale des sciences appliquées al Hoceima .....	3
Figure 2 Cycle de vie du projet data science .....	4
Figure 3 Trello Overview .....	8
Figure 4 Git .....	9
Figure 5 Questionnaire .....	14
Figure 6 Représentation des données .....	14
Figure 7 Les types de données .....	15
Figure 8 cas de données à nettoyer.....	18
Figure 9 transformation des données.....	18
Figure 10 transformer les données qualitatives en données quantitatives .....	19
Figure 11 la fonction get_dummies.....	19
Figure 12 Machine Learning Architecture.....	20
Figure 13 Matrice de Corrélation .....	22
Figure 14 Analyse descriptive.....	23
Figure 15 Diagramme de composantes .....	23
Figure 16 Diagramme sequence pour l'autentification .....	29
Figure 17 Diagramme de séquence pour la demande de prédiction.....	30
Figure 18 shema technique du projet.....	32
Figure 19 l'editeur spyder .....	38
Figure 20 Anaconda Prompt .....	39
Figure 21 Feuille Excel.....	41
Figure 22 Creation du dashboard sur Power Bi .....	42
Figure 23 L'editeur Sublime text.....	43
Figure 24 Page d'accueil .....	44
Figure 25 Page de prédiction .....	45
Figure 26 :L'etat psychique des étudiants dans l'examen .....	50
Figure 27 distrubition de réussite selon le sexe .....	51
Figure 28 taux de satisfaction des étudiants par le rythme de formation au sein de l'ensa hoceima .....	52
Figure 29 l'etat psychique des etudiants dans la semaine de préparation.....	52



# Introduction générale

Les données sont au cœur des discussions actuelles, leur collecte, leur propriété, leur traitement, leur utilisation, Elles représentent un enjeu majeur pour les entreprises et pour les cabinets d'expertise-comptable. Les données, de tout temps un sujet sensible.

Dans notre projet on vise à collecter les données des élèves ingénieurs de l'école nationale des sciences appliquées d'Al Hoceima en vue de mettre en évidence les facteurs explicatifs de la réussite universitaire en les différents niveaux ainsi que leurs interactions afin de pouvoir classer les étudiants selon leur performance

En nous basant sur ces données, nous avons retenu pour expliquer la performance académique aussi bien des facteurs processuels (c'est-à-dire évoluant au cours de l'année académique, comme sa motivation, sa perception de son environnement universitaire ou sa confiance en lui) que structurels (facteurs invariants tels que ses antécédents scolaires et familiaux ou son sexe). Durant ce semestre, nous avons réalisé et distribué un questionnaire auprès de notre école. Nous allons donc être à même de relever les correspondances et les divergences entre les modèles prédictifs obtenus dans des institutions universitaires conduisant à des diplômes à la fois proches et différents

L'échantillon de données utilisé pour obtenir les résultats qui sont présentés ici contient 500 étudiants de différentes filières inscrits dans l'école nationale des sciences appliquées d'Al Hoceima. Chaque étudiant est décrit au moyen de 30 variables relevées par le biais du questionnaire. Une sélection de variables sera donc nécessaire préalablement à un quelconque traitement statistique ou mathématique. La variable à expliquer utilisée pour la construction de nos modèles est une variable à trois modalités, construite a posteriori, regroupant les étudiants selon leur performance académique

Notre objectif final est de pouvoir classer les étudiants en trois groupes : le groupe des étudiants ayant une forte probabilité de réussir l'année (Low Risk), le groupe des étudiants qui peuvent éventuellement réussir moyennant des actions à mener (Medium Risk) et le groupe des étudiants ayant une forte probabilité d'échouer (ou d'abandonner) (High Risk).

Afin de cibler correctement les étudiants qui ont vraiment besoin de mesures d'accompagnement, notre objectif est donc bien de proposer une méthode permettant d'identifier les étudiants Medium Risk le plus tôt possible dans l'année, avant la première session d'examens et ce afin d'optimiser la répartition des ressources pédagogiques visant à lutter contre l'échec scolaire, le désintérêt des études.

Afin de proposer une démarche visant à construire un modèle de prédiction de la réussite universitaire, nous présenterons avant tout la méthodologie que nous avons adoptée. Ensuite la visualisation de ces données dont nous disposons pour rendre compréhensibles les données les plus importantes et ce qu'elles signifient, nous présenterons les différents résultats obtenus par les méthodes des arbres de décision et L'analyse en composants principales et enfin, nous comparerons leurs performances avec celles de l'analyse discriminante linéaire.

Le présent document se compose de trois chapitres détaillant toutes les phases de la réalisation de notre projet. Le premier chapitre envisage la description du contexte général du projet qui permettra de situer ce dernier et de bien comprendre la problématique tout en mettant en valeur les services qui peut offrir la plateforme. Dans le deuxième chapitre la réalisation de l'ensemble de phases du projet, le troisième chapitre on va aborder l'étude fonctionnelle du projet, qui vous donnera un aperçu sur les principales fonctionnalités de notre système. Puis on va passer au quatrième chapitre qui s'agit de l'étude technique du projet qui décrit ce dernier en termes d'architecture. Et finalement le cinquième chapitre décrit l'étape de la réalisation et mise en œuvre.

# Contexte Du Projet

Ce chapitre présente le contexte général du projet. Nous allons commencer tout d'abord par une présentation du projet où nous décrirons les objectifs ciblés ainsi que le planning de travail adopté pour ce projet.

## 1.1 Présentation de Projet

### 1.1.1 Présentation de l'école nationale des sciences appliquées d'Al Hoceima (ENSAH)

L'Ecole Nationale des Sciences Appliquées d'Al Hoceima (ENSAH) est un établissement public d'enseignement supérieur relevant de l'université Abdelmalek Essaadi. Créée en 2008, Sa création s'inscrit dans l'optique de favoriser la formation des ingénieurs d'Etat hautement qualifiés dans les spécialités les plus ouvertes et susceptibles de connaître d'importants développements au sein du tissu socio-économique régional et national.



*Figure 1 l'école nationale des sciences appliquées al Hoceima*

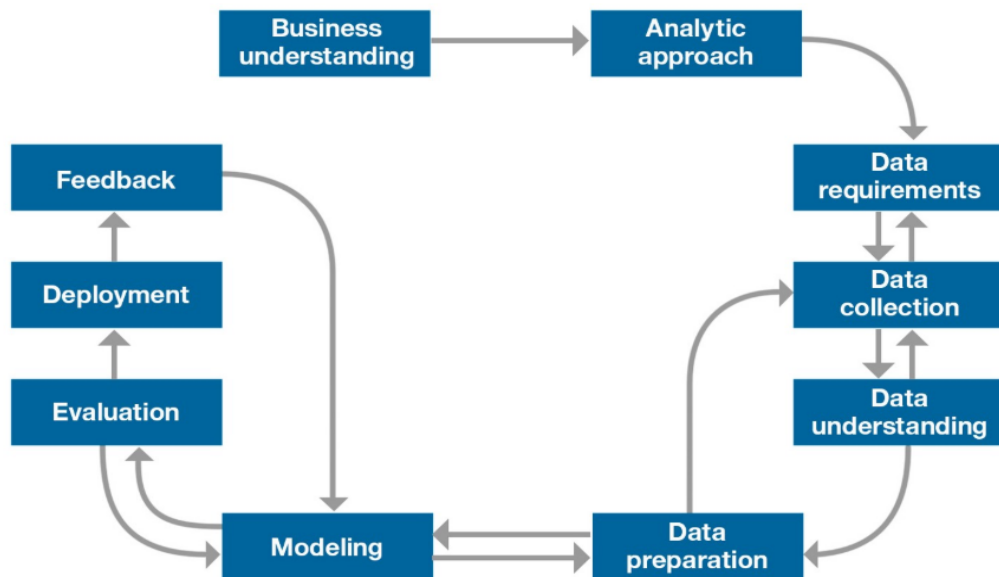
L'Ecole Nationale des Sciences Appliquées Al-Hoceima dispense des formations de deux années préparatoires (AP1, AP2) et trois années de spécialités dans les domaines suivants :

- Génie Informatique
- Génie Civil
- Génie Energétique et énergies renouvelables
- Génie de l'eau et Environnement

### 1.1.2 Notions de base

### 1.1.2.1 Data Science

La data science (Science de donnée) est le processus d'inspection, nettoyage, transformation et modélisation des données pour la prise de décisions et avoir des réponses à nos questions.



*Figure 2 Cycle de vie du projet data science*

### **1.1.2.2 La science de données au service de l'université**

La science des données (en anglais data science) est une discipline qui s'appuie sur des outils mathématiques, de statistiques, d'informatique (cette science est principalement une « science des données numériques » ) et de visualisation des données. Elle est en plein développement, dans le monde universitaire ainsi que dans le secteur privé et le secteur public.

La technologie de l'information (TI) est un pilier indispensable de la société moderne. Nous assistons maintenant à une révolution dans le domaine des technologies de l'information, au centre de laquelle figure «Big Data» et DEEP Learning, où nous générons, collectons, stockons et pouvons traiter des données à des taux phénoménalement plus élevés que ce qui est soutenable.

Grâce au Big Data et à l'essor de l'informatique et du numérique dans l'éducation, l'analyse des données permet désormais d'améliorer les systèmes scolaires et éducatifs du monde entier. Cette numérisation de l'éducation génère un très grand volume de données relatives à l'apprentissage et à l'enseignement. Les entreprises technologiques et les établissements scolaires peuvent dorénavant s'associer pour convertir ces données en pistes à suivre pour développer de meilleures méthodes d'enseignement de nouveaux programmes scolaires, et pour remédier aux problèmes des étudiants en difficulté. Voici plusieurs exemples concrets de l'utilisation de l'analyse des données dans le secteur de l'éducation :

- Obtenir une vue d'ensemble sur le fonctionnement des établissements.
- Améliorer les résultats et l'expérience des étudiants.
- Lutter contre l'échec scolaire.
- Développer l'apprentissage à distance.
- Améliorer l'enseignement supérieur

### **1.1.3 Problématique**

La formation tout au long de la vie intègre des expériences d'études variées, de la formation initiale traditionnelle à la reprise d'études à l'âge adulte pour des formations créditées ou non, courtes ou longues, à temps plein ou à temps partiel et dans le but d'une qualification professionnelle ou pour des intérêts personnels et culturels. Dans une perspective comparée et sur la base de l'indicateur commode de l'âge, les universités marocaines répondent mal à cette ambition.

Or, l'évaluation de cette contribution ne peut se poser de façon isolée sans comprendre que l'éducation des adultes a été problématisée dans l'espace public national en dehors des périmètres de compétences et des catégories de perception et d'intervention de la politique universitaire.



Dans les Universités, les cours magistraux sont par nature moins interactifs que dans les niveaux inférieurs d'enseignement. Par conséquent, les enseignants reçoivent peu de retours sur l'efficacité de leurs méthodes, jusqu'à ce que les étudiants réussissent ou échouent à leurs examens.

Face à cette situation critique dans les universités et les grandes écoles, et en parallèle avec ce qu'on a dit dans la partie précédente « la science des données au service de l'université » Notre objectif principal durant cette étude est de préciser les causes et définir tous les acteurs directs et indirects qui influencent la vie scolaire d'un étudiant en s'appuyant sur les différentes technologies disponibles surtout que dans le passé, le principal symptôme d'un échec scolaire était une baisse des résultats aux examens. Mais aujourd'hui, grâce à l'analyse continue des données de chaque élève, il est possible de proposer à chacun un apprentissage plus personnalisé en prenant en compte les centres d'intérêt, les connaissances personnelles et les capacités intellectuelles individuelles.

### 1.1.4 Contexte du Projet

L'objectif est de réaliser une plateforme de prédiction basée sur l'analyse des données des étudiants des postes est de tirer les sujets les plus abordés par un prol donné an d'avoir une idée sur les centres d'intérêt de ce projet, pour se faire on a suivi ces étapes :

1. **L'extraction des données** : Cette étape consiste à collecter les informations de la population cible en s'intéressant au .....
2. **Structuration des données** : après l'extraction on vient à la structuration sous une format plus facile et convenable à l'utilisation.
3. **Conception des algorithmes du traitement des données** : Cette étape est le coeur de notre application, elle se compose des algorithmes statistiques et d'autres pour la classification de ces données pour en tirer les informations voulues.
4. **Visualisation des données** : pour rendre compréhensible et mettre en valeur les données traitées.

## 1.2 Conduite et gestion de Projet

La gestion de projet est une démarche visant à organiser de bout en bout le bon déroulement d'un projet. Lorsque la gestion de projet porte sur un ensemble de projets concourant à un même objectif, on parle de gestion de programme.

En pratique, le projet est tourné vers l'objectif final, il doit être adaptable à des modifications fréquentes, mais maîtrisé et planifié. Donc toute modification doit rester planifiée. Et notamment, le projet doit rester dynamique et équilibrer continuellement les contraintes techniques, de coût et de délai.

### 1.2.1 Contraintes de Projet

Cette section recense les contraintes ou les restrictions particulières qui limitent le projet :

#### **Contraintes techniques**

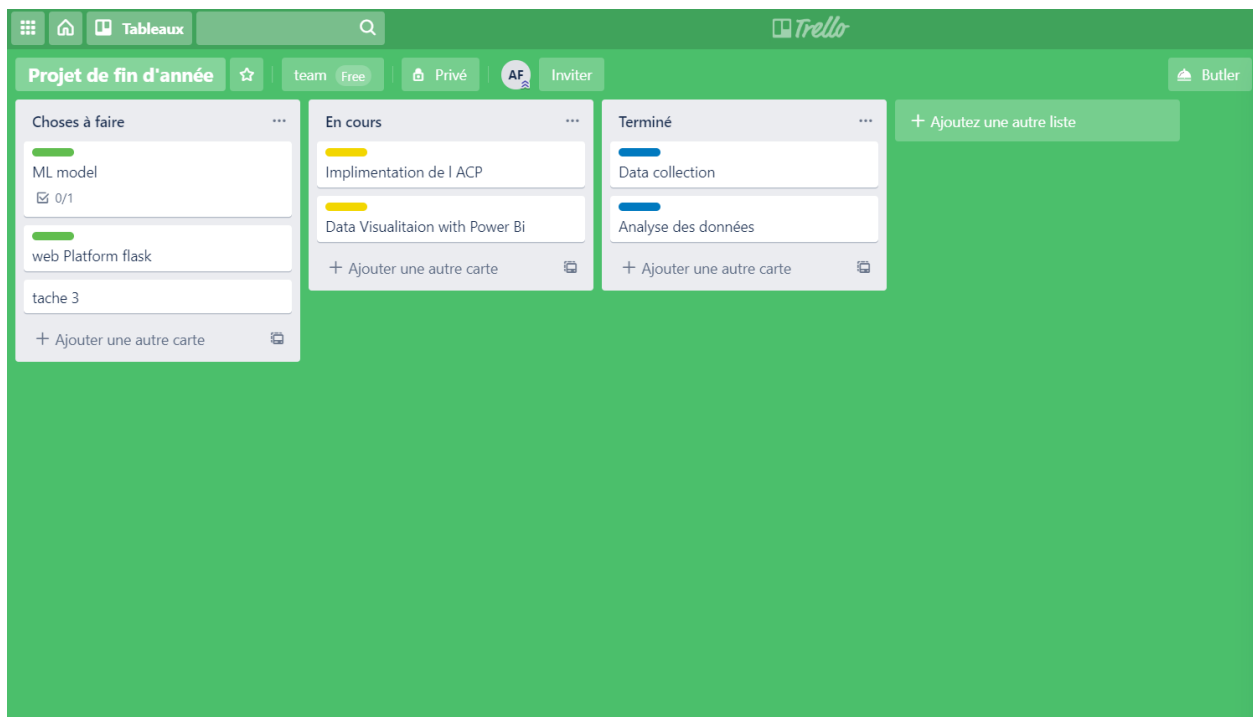
1. *La qualité de données* : Un algorithme basique utilisant des données de qualité sera généralement beaucoup plus performant qu'un algorithme généralement très performant mais avec des données moins pertinentes ou de moins bonne qualité.

L'efficacité d'un modèle prédictive nécessite des données de qualité pour avoir des bons résultats, pour ce faire, il faut traiter divers information sur les étudiants à voir leurs notes académiques.

2. *Le suivi d'étudiant* : Le suivi du parcours des étudiants durant les cinq années de formation dans l'école nationale des sciences appliquées pour bien avoir une étude plus efficiente.

## 1.2.2 Gestion de Projet avec Trello

Trello est un outil en ligne pour la gestion de projets et de tâches personnelles. Cela peut sembler plutôt prosaïque. Mais cette application de plus en plus populaire inspire souvent le genre de passion habituellement réservé aux applications grand public comme Pinterest ou Instagram. C'est le genre de logiciel d'entreprise qui se glisse dans les entreprises à travers la porte dérobée, juste parce que les employés individuels aiment comment cela fonctionne. Trello est la manière simple, gratuite, flexible et visuelle de gérer vos projets et d'organiser tout ce qui est digne de confiance de millions de personnes de partout dans le monde.



*Figure 3 Trello Overview*

Trello vient avec plusieurs valeurs ajoutées citons :

- Ouvrir une carte et ajouter des commentaires, télécharger des pièces jointes, créer des listes de vérification, ajouter des étiquettes et des dates d'échéance, et plus encore.
- Commencer une discussion avec des commentaires et des pièces jointes. Ajoutez une note rapide ou une diatribe détaillée, validée par Markdown.
- Mentionner un membre dans un commentaire pour s'assurer qu'il est averti. Ajouter des fichiers en les téléchargeant depuis l'ordinateur, Google Drive, Dropbox, Box et OneDrive. L'égayer avec des emojis, aussi
- Chaque fois que quelque chose d'important arrive, vous savez instantanément avec le système de notification de Trello. Vous recevez des notifications où que vous soyez : à l'intérieur de l'application, par e-mail, via des notifications sur le bureau via le navigateur ou via des notifications push mobiles. Les notifications restent synchronisées sur tous vos appareils. Dans notre cas on utilise les notifications avec Slack de façon que nous somme notifiés pour toute modification dans le tableau de projet.
- Les données sont privées et sécurisées. Vous avez un contrôle total sur qui voit vos tableaux. Toutes les données sont envoyées via une connexion SSL / HTTPS sécurisée, la même technologie de cryptage que celle utilisée par les banques.

### 1.2.3 Gestion des versions avec Git

A cause de la pandémie de Covid-19 nous étions obligés de poursuivre ce travail à distance et en équipe, chose qui nous a poussés à chercher une solution pour que chacun puisse travailler sur une version de code source sans modifier le code source existant, la solution est Le Contrôle de Version.



*Figure 4 Git*

Les systèmes de contrôle de version sont une catégorie d'outils de développement qui aident une équipe de développement à gérer les changements apportés au code source au fil du temps. Les logiciels de contrôle de version gardent une trace de chaque changement apporté au code dans un type spécial de base de données. Si une erreur est commise, les développeurs peuvent revenir en arrière et comparer les versions antérieures du code, ce qui leur permet de corriger l'erreur tout en minimisant les perturbations pour tous les membres de l'équipe.

Git est de loin le système de contrôle de version le plus largement utilisé aujourd'hui. Git est un projet open source avancé, qui est activement maintenu. À l'origine, il a été développé en 2005 par Linus Torvalds. De plus en plus de projets de développement reposent sur Git pour le contrôle de version, y compris des projets commerciaux et en open source. Les développeurs qui utilisent Git sont bien représentés dans le pool de talents disponible, et la solution fonctionne bien sur une vaste gamme de systèmes d'exploitation et d'environnements de développement intégrés (IDE).

En plus d'être décentralisé, Git a été conçu pour répondre à trois objectifs : performances, sécurité et flexibilité.

- Les performances brutes de Git sont très élevées par rapport à de nombreuses alternatives. Commuter de nouveaux changements, créer des branches, faire des merges et comparer les anciennes versions.
- L'intégrité du code source géré a été la priorité absolue lors de la conception de Git. Le contenu des fichiers, les liens entre les fichiers et les répertoires, les versions, les tags, les commits : tous ces éléments du dépôt Git sont sécurisés à l'aide d'un algorithme de hachage sécurisé de façon cryptographique, appelé SHA1. Celui-ci protège le code et l'historique des changements contre toute modification accidentelle ou malveillante, tout en assurant une traçabilité complète de l'historique, créer des branches, faire des merges et comparer les anciennes versions.
- Lors de la conception de Git, l'un des principaux objectifs a été la flexibilité. Git est flexible à plusieurs niveaux : que ce soit pour la prise en charge de divers workflows de développement non linéaires, pour son efficacité dans l'élaboration de projets de différente envergure ou pour sa compatibilité avec de nombreux systèmes et protocoles existants.

## **1.3 Conclusion**

Dans le présent chapitre, nous avons situé notre projet de fin d'année, en présentant d'une part le contexte général et la problématique, ainsi que les concepts clés de notre sujet. D'autre part nous avons présenté l'enchaînement et le déroulement de notre projet.

# Le Cycle De Vie Du Projet

On va développer dans ce chapitre les principaux axes de connaissance ci-dessous sur lesquels nous avons travaillé pour structurer nos recherches et nécessaires pour mieux comprendre le projet et ses enjeux techniques :

- La collecte de données
- La préparation des données
- Présentation et domaines d'application du machine learning
- Les différents types d'apprentissage
- Les modèles générés par les algorithmes d'apprentissage et leur évaluation
- Les architectures typiques de la machine Learning
- Une proposition de cycle de vie efficace pour un projet de machine Learning

## 2.1 La Collecte De Données

Dans notre projet on vise à collecter les données via une formulaire distribué aux élèves ingénieurs de l'école nationale des sciences appliquées d'Al Hoceima en vue de mettre en évidence les facteurs explicatifs de la réussite universitaire en les différents niveaux ainsi que leurs interactions afin de pouvoir classer les étudiants selon leurs performances.

L'échantillon de données utilisé pour obtenir les résultats qui sont présentés ici contient 500 étudiants de différentes filières inscrits dans l'école nationale des sciences appliquées d'Al Hoceima. Chaque étudiant est décrit au moyen de 30 variables relevées par le biais du questionnaire. Une sélection de variables sera donc nécessaire préalablement à un quelconque traitement statistique ou mathématique. La variable à expliquer utilisée pour la construction de nos modèles est une variable à trois modalités, construite a posteriori, regroupant les étudiants selon leur performance académique.

Lesquels , parmi les facteurs suivants , ont influencés negativement l'interet a vos etudes ? \*

	Oui	Non
Services bibliothèques	<input type="radio"/>	<input type="radio"/>
Transport Scolaire	<input type="radio"/>	<input type="radio"/>
Orientation	<input type="radio"/>	<input type="radio"/>
ressources materials des travaux	<input type="radio"/>	<input type="radio"/>
Mauvais Gestion de temps	<input type="radio"/>	<input type="radio"/>
Qualité de logement	<input type="radio"/>	<input type="radio"/>
l'addiction aux drogues	<input type="radio"/>	<input type="radio"/>
le mauvais choix des cours	<input type="radio"/>	<input type="radio"/>
Les réseaux sociaux	<input type="radio"/>	<input type="radio"/>
Barrières relies au langage	<input type="radio"/>	<input type="radio"/>
l'entourage	<input type="radio"/>	<input type="radio"/>
Gaming	<input type="radio"/>	<input type="radio"/>



### 2.1.1 Représentation d'échantillon

On représente en général un échantillon sous forme de tableau, où chaque ligne correspond à un individu, et chaque colonne représente une variable. Cette représentation est à l'origine du format de fichier CSV (comma separated values). Ce format est facilement interprétable par Python Avec Jupyter Nootbook

```
#On import notre dataset
pathfile = "C:/Users/Avicii/Desktop/Samad/GI2/Projet de fin d'etude/PredictiveModel/data/data.xlsx"
df = pd.read_excel(pathfile)
df.head()
```

sexe	age	logement	filere	endroit_travail	travail_equipe	heure_dormir	sport_par_sem	activites_para	...	ressources_materials	Mauvais_Gestion_temps	q
Homme	21	Interne	GI	Cafè	Non	00:00:00	2 fois	oui	...	Non	Oui	
Homme	18	Interne	CP1	La Bibliothèque	Non	02:00:00	5	Oui	...	Oui	Oui	
Homme	21	Interne	GI	Cafè	Oui	01:00:00	3 fois	Oui	...	Oui	Oui	
Homme	18	Externe	CP2	Cafè	Oui	02:00:00	5	Oui	...	Non	Non	
Homme	21	Interne	GC	Chez moi	Oui	01:00:00	3 fois	Non	...	Oui	Non	

JMns

Figure 6 Représentation des données

## 2.2 Compréhension Des Données

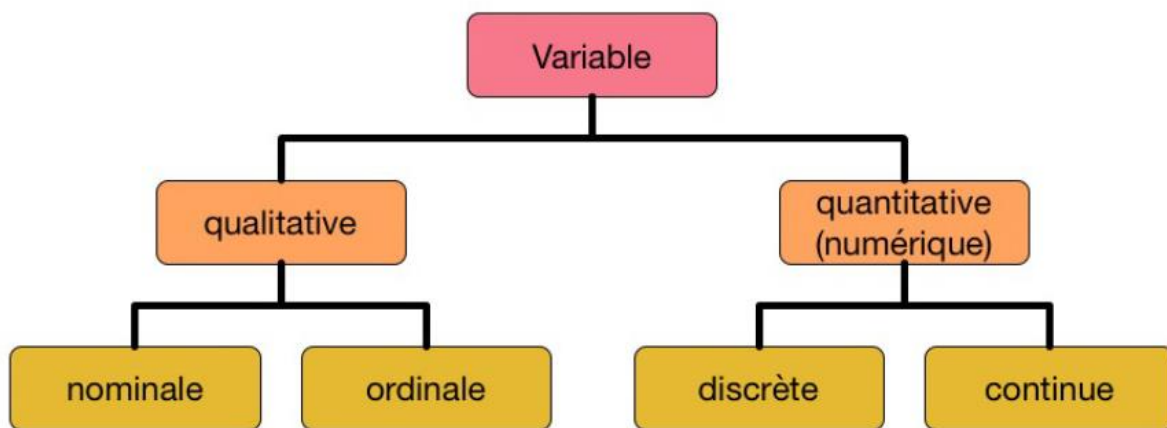
Lorsque les données ont été collectées, il est important de découvrir les différents data sets, ce qui permet de mieux connaître les données et de déterminer le traitement à leur appliquer avant qu'elles deviennent exploitables dans l'étape de modélisation.

## 2.2.1 Les Types De Données

Une variable est une caractéristique pouvant prendre plusieurs des valeurs d'un ensemble d'observations possibles, auquel une mesure ou une qualité peut être appliquée. Par exemple, la variable « Sexe » peut avoir 2 observations différentes possibles : « Masculin » et « Féminin ». D'autres variables possibles sont l'âge, la filière, l'espace préféré d'étude, etc.

La distinction entre les types de variables décide des analyses statistiques utilisables, des représentations graphiques utilisables ou encore des tests statistiques. On doit pouvoir choisir les bonnes analyses pour que le travail soit valide.

Il y a deux types de variables, chacun d'eux est subdivisé en deux groupes :



*Figure 7 Les types de données*

### 2.2.1.1 Les Variables Qualitatives

Comme le nom l'indique, les variables qualitatives contiennent des valeurs qui expriment une qualité comme le sexe, la couleur ou le nom. Elles peuvent être :

- **Des variables qualitatives Nominale** : Une variable est qualitative nominale quand ses valeurs sont des éléments d'une catégorie non hiérarchique. C'est-à-dire que ses éléments ne peuvent pas se ranger dans une gradation logique.

Exemple :

- ✓ Sexe : Masculin, Féminin
- ✓ Lieu préféré de travail : Café, la bibliothèque, Chez moi.

- **Des variables qualitatives Ordinale :** Une variable est qualitative ordinale quand ses valeurs sont des éléments d'une catégorie hiérarchique. C'est-à-dire que ses éléments peuvent être rangés dans une gradation logique.

Exemple :

- ✓ Niveau d'étude : Cycle préparatoire, Cycle d'ingénieur
- ✓ Activités sportives : Jamais, rarement, à l'occasion, fréquemment, toujours.
- ✓ Cas psychique dans l'examen : Stressé, Pas mal, Normal

On a ainsi une gradation des valeurs, une relation d'ordre entre chacune des appréciations.

### 2.2.1.2 Les Variables Quantitatives

Les données ou variables quantitatives contiennent des valeurs numériques faisant référence à une unité de mesure reconnue. Pour cette raison, elles sont quelques fois qualifiées de variables métriques. L'âge, le nombre de modules rattrapage dans un semestre ou bien encore la population (dans le sens du nombre des étudiants) sont des variables quantitatives.

Ce sont les variables qui prennent des valeurs numériques, à condition que ces valeurs expriment une quantité et aient un sens lorsque l'on y applique des opérations arithmétiques.

Si le nombre de valeurs possibles (et probables) d'une variable est très grand, alors on peut la considérer comme *continue*. Sinon, on la considère comme *discrète*

## 2.3 Préparer Les Données

Cette phase de préparation des données intervient une première fois avant la phase de modélisation mais pourra être affinée plusieurs fois pendant la modélisation. Elle sera ensuite automatisée. La préparation des données est estimée par de nombreux experts comme étant souvent la phase la plus longue et la plus critique d'un projet de machine Learning. Cette phase peut s'organiser comme suit :

1. Sélectionner les données retenues : toute donnée non pertinente ou redondante doit être éliminée pour éviter d'augmenter la complexité du problème et le temps de calcul inutilement.
2. Nettoyer les données : il y a par exemple couramment des données erronées ou des cas d'exception inutiles comme exemple à utiliser pour l'apprentissage.
3. Enrichir les données : des données mieux construites sur des données existantes permettront souvent de grandement améliorer la performance du ML.
4. Intégrer les données des différentes sources : la plupart des algorithmes ML apprennent avec des exemples qui doivent contenir toutes les données agrégées à chaque exemple sous la forme suivante : "toutes les données attendues + résultat correspondant"
5. Standardisation des données (moyenne, plage, catégorisation, imputation, normalisation, binarisation) : différentes options de standardisation des données peuvent être requises pour améliorer la performance de l'algorithme ou pour être compatible. Par exemple, beaucoup d'algorithmes ML sont plus performants si la moyenne des données est centrée sur zéro ou que les données sont comprises entre 0 et 1. Certaines données peuvent être manquantes dans de nombreux exemples, il faut prévoir une stratégie adaptée (ex. imputation). Chaque algorithme accepte aussi un certain nombre de formats, il est donc souvent nécessaire de reformater les données au préalable.
6. Créer différents jeux de données séparés pour : l'apprentissage, le test, la validation. Il est courant d'utiliser le système de validation croisée pour les jeux d'apprentissage et de test (voir chapitre suivant) mais il existe aussi d'autres stratégies.

### 2.3.1 Nettoyage et transformation Des Données

Dans la figure au-dessous on a des exemples pour une valeur manquantes et une valeur aberrante

F	G	H	I
ton meilleur endroit d'étudier . ?	Vous travaillez en équipe	A quelle	Combien de fois par semaine vous faites de sport ?
Chez moi	Non	01:00:0	jamais
Chez moi	Non	05:00:0	jamais
Chez moi	Non	02:30:0	3 fois
Chez moi	Non	00:30:0	jamais
Chez moi	Non	01:00:0	jamais
Cafè	Oui	02:00:0	Kena w Mab9inaash akhoya
Cafè	Oui	02:00:0	Occasionnellement
Cafè	Non	01:00:0	2 fois
	Non	03:00:0	une fois

Figure 8 cas de données à nettoyer

Pour les variables qualitatives nous utilisons les étiquettes de colonne et de ligne pour accéder aux données avec .loc du package Pandas.

```
#on change l'ancienne valeur par "jamais"  
df.loc[df["sport_par_sem"] == "kenna w mab9inash a khoya", "sport_par_sem"] = "jamais"
```

Figure 9 transformation des données

### Transformation des données

Comme les algorithmes de machine Learning ne peuvent traiter que les variables numériques alors il est nécessaire de transformer les données catégoriques en valeurs numériques

```
df = df.replace(to_replace="Oui", value=1)
df = df.replace(to_replace="Non", value=0)
df.head()
```

*Figure 10 transformer les données qualitatives en données quantitatives*

A l'aide de la fonction `.get_dummies` du package Pandas on peut convertir simplement nos variables catégoriques. Dans cet exemple on a transformer la variables sexe en affectant 0 à « Homme » et 1 à « Femme » en vue de rendre les données plus simples à traiter avec les algorithmes d'apprentissage automatiques.

```
dummy_variable_1 = pd.get_dummies(df["sexe"])
dummy_variable_1.head()
```

	Femme	Homme
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1

*Figure 11 la fonction `get_dummies`*

## 2.4 Le « Machine Learning » : Connaissances nécessaires

La machine Learning (ML), ou l'apprentissage automatisé, est un champ de l'intelligence artificielle. Il peut être défini comme couvrant l'étude, la conception et le développement d'algorithmes donnant la possibilité à des machines d'apprendre sans avoir été explicitement programmées (définition d'Arthur Samuel en 1959). un algorithme de ML va analyser de nombreux cas d'exemple pour produire un programme ou modèle pour effectuer la tâche illustrée par les exemples. Un tel algorithme peut combiner un très grand nombre de données et règles différentes, et être mis à jour en lui fournissant de nouveaux exemples.

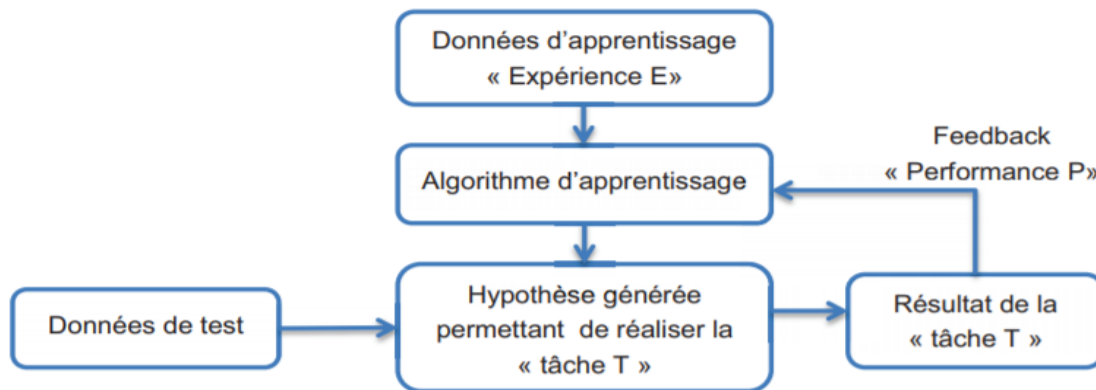
Une tâche de machine Learning peut ainsi apprendre à : classifier, prédire, recommander, optimiser, détecter des motifs ou anomalies, filtrer... Les tâches de machine Learning peuvent évidemment être combinées avec d'autres algorithmes ML ou non pour produire des algorithmes plus avancés. Dans notre cas on cherche à classifier les étudiants en trois classes (High Risk, Medium Risk, Low Risk) selon leurs parcours universitaire.

### 2.4.1 Objectif et caractéristiques du problème d'apprentissage

Pour présenter le domaine, nous commencerons par une citation (Mitchell 1997) de Tom M. Mitchell, directeur du département de Machine Learning à la Carnegie Mellon University.

Il s'agit d'améliorer la performance  $P$  d'une machine à réaliser la tâche  $T$  en utilisant un ensemble d'exercices  $E$

On dit qu'une machine apprend si : sa performance sur une tâche de  $T$  mesurée par  $P$  augmente avec l'expérience  $E$ .



*Figure 12 Machine Learning Architecture*

Dans notre cas :

- L'expérience sera nos données d'apprentissage souvent exemples.
- L'hypothèse générée par l'algorithme d'apprentissage sera appelée « Modèle »
- Les mesures de performances sont couramment : la précision, le taux d'erreur, la variance entre le résultat donné par l'algorithme et le résultat attendu.

Un problème d'apprentissage peut ainsi se caractériser par :

- Un type d'apprentissage définissant la façon d'interagir avec l'environnement.
- Une sortie dont on va mesurer l'erreur généralement sous d'une forme d'une « fonction de cout » à minimiser.
- Un « modèle » et ses paramètres.
- Un algorithme pour créer et adopter le modèle en utilisant les exemples d'apprentissage issus de l'environnement, de façon à optimiser la fonction de cout.

## 2.4.2 Types d'apprentissage

### 2.4.2.1 Apprentissage supervisé

Dans l'apprentissage Supervisé on dispose d'un ensemble d'objets et pour chaque objet une valeur cible associée, il faut apprendre un modèle capable de prédire la bonne valeur cible d'un objet nouveau.

Il existe deux types de sous-problèmes en apprentissage supervisé numérique :

1. **Régression** ("Régression") : lorsque la valeur cible à prédire est continue.
2. **Classification** ("Classification") : lorsque la valeur cible à prédire est discrète

Dans le cadre de notre étude on cherche à classer les étudiants en 3 classes (High Risk, Medium Risk, Low Risk) en intégrant les algorithmes de régression et de classification.



### 2.4.2.2 Apprentissage non supervisé :

Dans l'apprentissage non supervisé on dispose d'un ensemble d'objets sans aucune valeur cible associée, il faut apprendre un modèle capable d'extraire les régularités présentes au sein des objets pour mieux visualiser ou appréhender la structure de l'ensemble des données.

### Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales (ACP) est l'une des méthodes d'analyse de données multi variées les plus utilisées. Elle permet d'explorer des jeux de données multidimensionnels constitués de variables quantitatives. C'est une méthode de projection qui permet de projeter les observations depuis l'espace à  $p$  dimensions des  $p$  variables vers un espace à  $k$  dimensions ( $k < p$ ) tel qu'un maximum d'information soit conservée (l'information est ici mesurée au travers de la variance totale du nuage de points) sur les premières dimensions. Si l'information associée aux 2 ou 3 premiers axes représente un pourcentage suffisant de la variabilité totale du nuage de points, on pourra représenter les observations sur un graphique à 2 ou 3 dimensions, facilitant ainsi grandement l'interprétation.

Dans notre cas on cherche à appliquer l'ACP pour :

L'étude et la visualisation des corrélations entre les variables pour limiter éventuellement le nombre de variables à mesurer par la suite

- La visualisation des observations dans un espace à deux dimensions, afin d'identifier des groupes homogènes d'observations.

En Appliquant l'analyse des composantes principales sur notre jeu de données on trouve les résultats suivants :

#### 1) La Matrice de Corrélation

Une matrice des corrélations permet de détecter rapidement certaines liaisons. C'est donc en amont des études qu'on l'utilise. Toutefois, dès que le nombre de variables devient important, les interprétations deviennent difficiles et on se tourne souvent vers les analyses factorielles.

On constate Parfois une indépendance linéaire presque totale entre certaines variables tandis qu'ils montrent autrefois des similitudes tantôt avec l'un, tantôt avec l'autre.

Matrice de corrélation <sup>a</sup>										
	travail_equipe	services_bibli othèques	transport_sco laire	orientation	ressources_ materials	Mauvais_Gest ion_temps	qualite_loge ment	mauvais_choi x_de_cours	réseaux_soci aux	Barrieres_de _langage
travail_equipe	1,000	-,008	,064	-,036	,068	-,110	-,002	,039	-,028	,015
services_bibliothèques	-,008	1,000	,148	,176	,119	,038	,151	-,004	-,061	,031
transport_scolaire	,064	,148	1,000	,166	,173	-,126	,016	-,059	-,050	,085
orientation	-,036	,176	,166	1,000	,242	,016	,093	,151	,014	,146
ressources_materials	,068	,119	,173	,242	1,000	-,065	,151	,128	-,014	,061
Mauvais_Gestion_temps	-,110	,038	-,126	,016	-,065	1,000	-,005	,094	,155	,023
qualite_logement	-,002	,151	,016	,093	,151	-,005	1,000	,191	,006	,178
mauvais_choix_de_cours	,039	-,004	-,059	,151	,128	,094	,191	1,000	,061	,171
réseaux_sociaux	-,028	-,061	-,050	,014	-,014	,155	,006	,061	1,000	-,008
Barrieres_de_langage	,015	,031	,085	,146	,061	,023	,178	,171	-,008	1,000

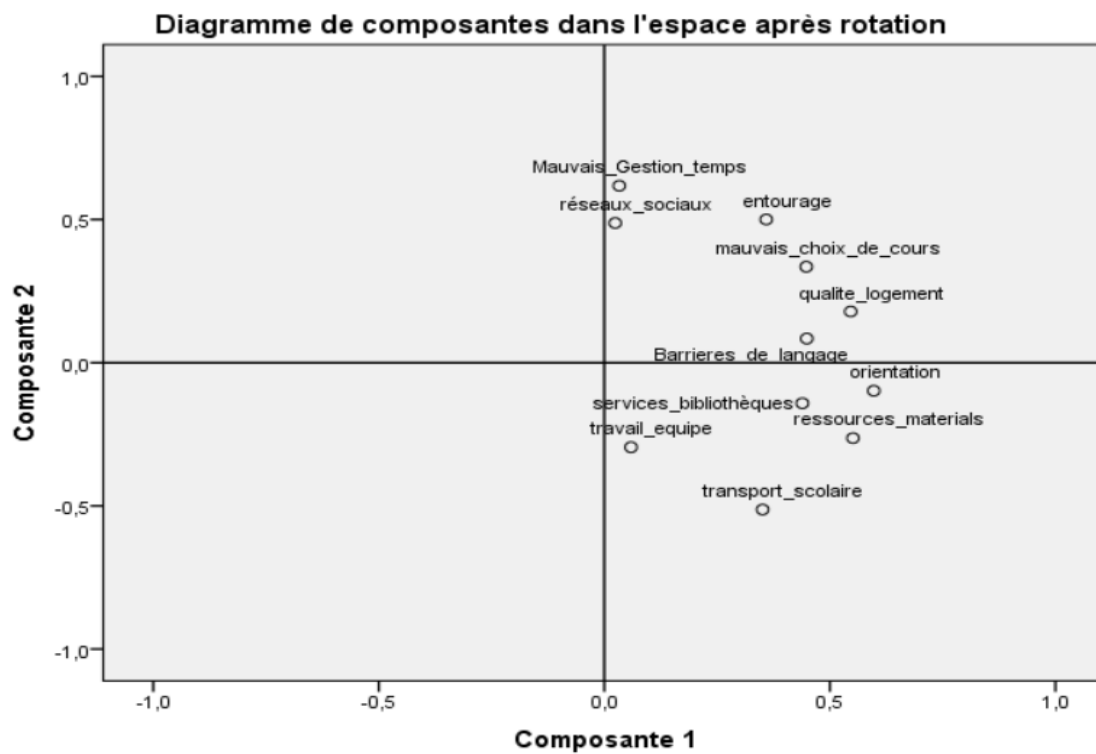
Figure 13 Matrice de Corrélation

## 2) L'analyse descriptive

Statistiques descriptives			
	Moyenne	Ecart-type	n analyse
travail_equipe	,39	,487	402
services_bibliothèques	,24	,425	402
transport_scolaire	,18	,386	402
orientation	,40	,491	402
ressources_materials	,56	,497	402
Mauvais_Gestion_temps	,84	,369	402
qualite_logement	,52	,500	402
mauvais_choix_de_cours	,58	,494	402
réseaux_sociaux	,56	,497	402
Barrieres_de_langage	,30	,459	402
entourage	,51	,500	402

*Figure 14 Analyse descriptive*

## 3) Le diagramme de composantes dans l'espace



*Figure 15 Diagramme de composantes*

D'après la carte factorielle on remarque qu'il y a probablement trois groupes.

Par exemple les deux variables « Mauvais Gestion du temps », « réseaux sociaux » sont bien corrélées ce qui explique la relation entre ces deux facteurs, car un étudiant qui passe pas mal de temps en navigant sur le réseaux sociaux aura sans doute un problème de gestion du temps et en fin un faible rendement dans ces études.

## **2.5 Conclusion**

Durant ce chapitre nous avons vus les différentes étapes de collecte et de préparation des données, cette dernière étape représente la partie la plus importante dans le cycle de vie projet avant d'entrer dans la phase de la modélisation ou il devient alors possible de former en continu le modèle et d'exploiter les résultats en apprenant à partir des données.

# Étude Fonctionnelle

Ce chapitre est consacré à la description et l'étude fonctionnelle qui concerne la traduction du besoin fonctionnel en fonctions de services à traiter, et tester chaque fonction pour objectif d'améliorer son fonctionnement ou ses propriétés et d'atteindre les objectifs fixés.

## 3.1 Analyse Fonctionnelle

L'étude ou l'analyse fonctionnelle est une démarche qui « *consiste à rechercher et à caractériser les fonctions offertes par un produit pour satisfaire les besoins de son utilisateur.* »

Cette démarche est une méthode d'analyse qui permet de transcrire le besoin d'un client d'une manière particulièrement précise et structurée, sans évoquer un moyen, mais tout en restant focalisée sur les détails de l'objectif.

Notre objectif de l'analyse fonctionnelle est de convertir le besoin client en fonctionnalités à remplir, identifie dans un premier temps les modules qui présentent l'ensemble des services du système, puis étudier dans un deuxième temps les différents diagrammes UML. Ce qui permet de sélectionner la meilleure réponse pour traiter le besoin client au mieux, tout en restant dans les clous du coût de développement.

### 3.1.1 Définir les besoins

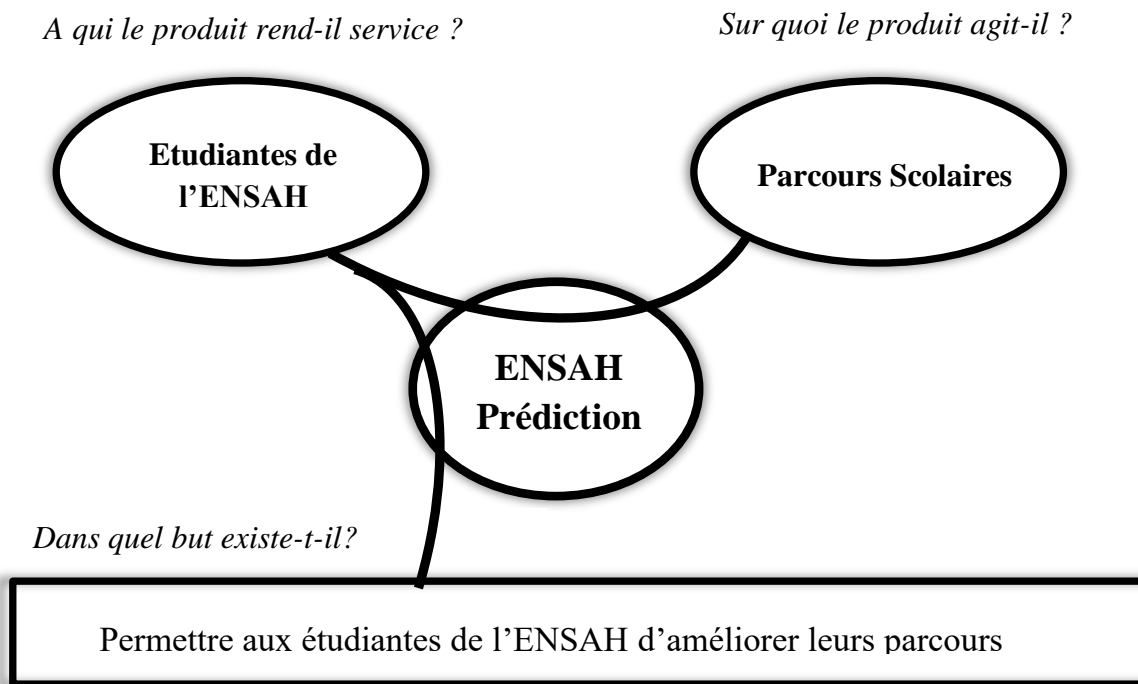
Avant de se lancer dans la conception, il convient de bien identifier et formaliser les objectifs du projet. Préciser les différents facteurs et modules. Cette phase est essentielle, car elle fixe la direction du travail qui va être entrepris.

Pour faciliter cette tâche, il existe des outils pratiques afin d'explicitier les besoins plus aisément.

#### 3.1.1.1 La Bête à Cornes

La bête à Cornes : Un diagramme bête à cornes est un outil pour l'analyse fonctionnelle du besoin. C'est un schéma qui démontre si le produit est utile pour l'utilisateur, s'il répond à ses besoins. C'est la première étape de la méthode APTE.

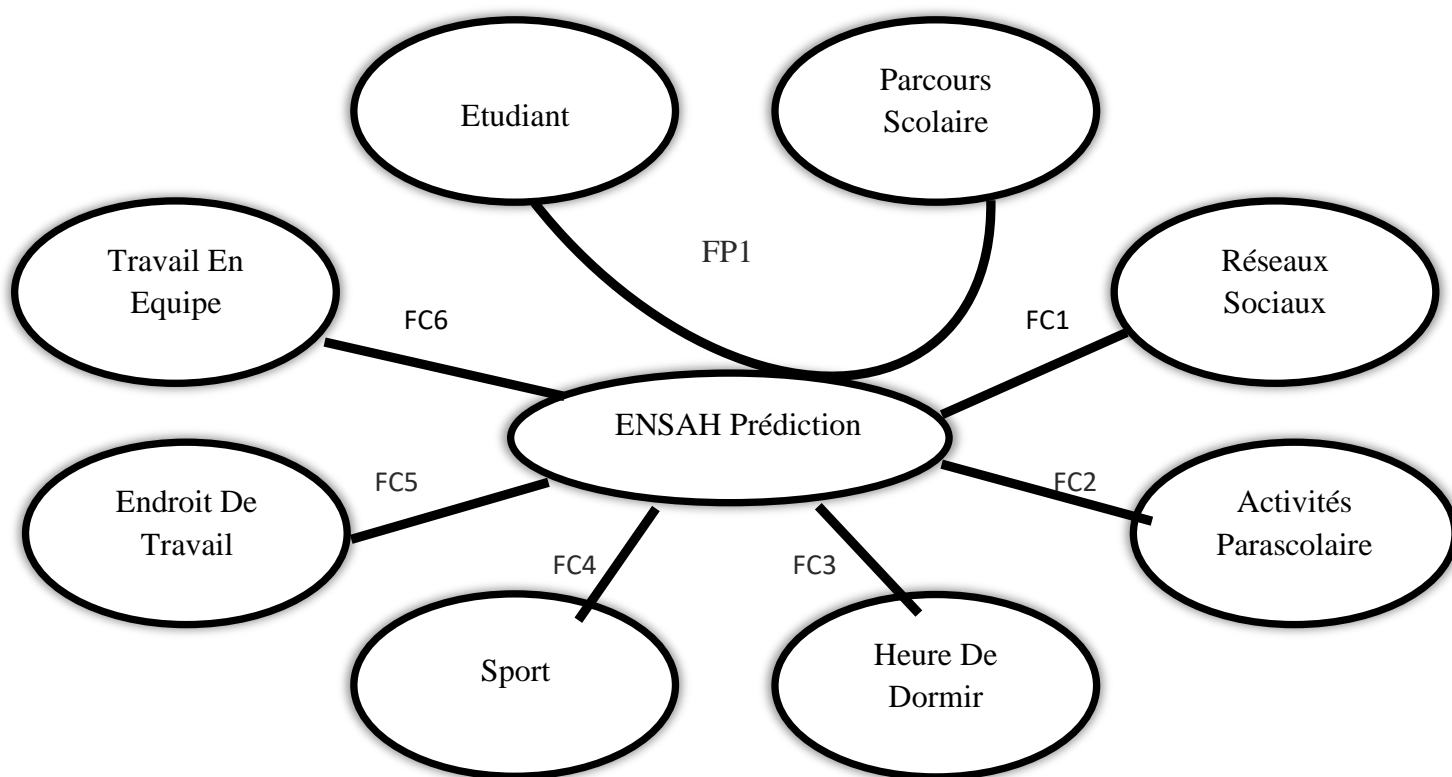
- *A qui rend-il service ?* : Les élèves ingénieurs de l'ENSA Hoceima
- *Sur quoi agit-il ?* : Le parcours scolaire durant toute cinq ans de formation
- *Dans quel but ?* Ou "pour quoi faire ?" : Permettre aux étudiantes de l'ENSAH d'améliorer leurs parcours scolaires



### 3.1.1.2 Diagramme Pieuvre

Diagramme Pieuvre : Le diagramme pieuvre ou graphe des interactions est un schéma qui représente la relation entre un produit/service et son environnement. C'est un outil d'analyse utilisé dans le cadre de la méthode APTE (Application aux Techniques d'Entreprise).

Le diagramme pieuvre permet de représenter les fonctions de service d'un produit. C'est-à-dire qu'il permet de voir quelles sont les fonctions essentielles et secondaires d'un produit et comment ces fonctions réagissent avec le milieu extérieur.



<b>Fonction</b>	<b>Description de la fonction</b>
FP1	Permettre aux étudiantes de l'ENSAH d'améliorer leurs parcours scolaires
FC1	Définir le nombre d'heures passé sur les réseaux sociaux
FC2	Suivre les activités parascolaires de l'étudiant
FC3	Savoir l'heure de dormir moyenne de l'étudiant
FC4	Suivre le nombre de jours de sport par semaine
FC5	Définir l'endroit de travail préféré pour l'étudiant
FC6	Suivre le travail de l'étudiant en équipe

### **3.1.2 Identifications des différents modules**

Le projet final est la collaboration de plusieurs applications afin de fournir une fonctionnalité majeure qui se résume en extraction d'informations personnelles et classification des données pour suggérer des recommandations pour améliorer votre parcours scolaire dans l'ENSAH.

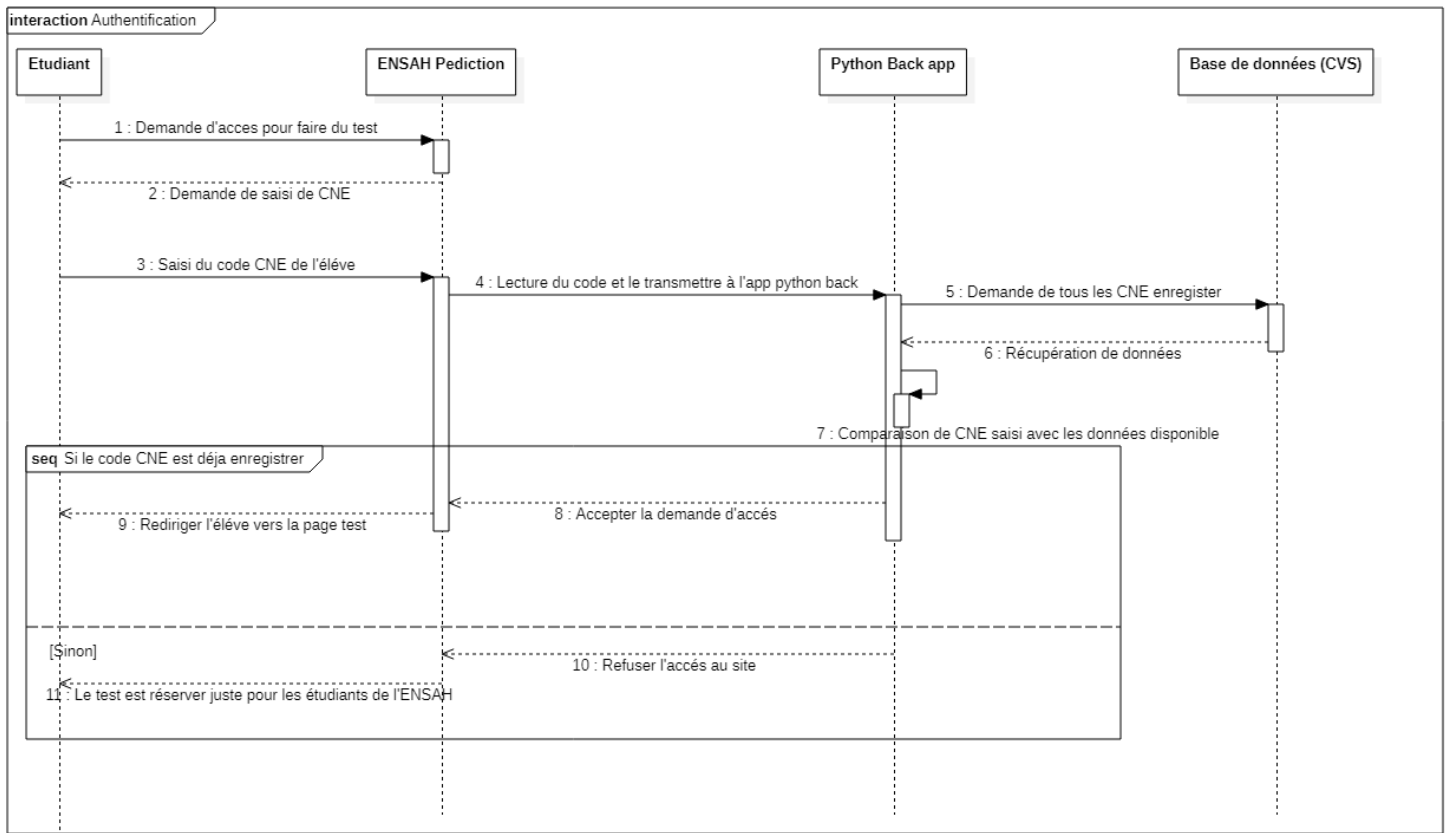
Pour un visiteur de la plateforme :

- \_ Authentifications avec le CNE de l'étudiant de l'EnsaH en utilisant
- \_ Remplissage de formulaire des informations personnelles.
- \_ Signaler un mot
- \_ Demander de modifier un mot
- \_ Ajouter une description pour un mot
- \_ Contacter les développeurs

### **3.1.3 Diagrammes de Séquence**

Les diagrammes de séquences peuvent servir à illustrer un cas d'utilisation

A cette étape il faut bien préciser les rôles entre les acteurs et le système selon un ordre chronologique bien définie pour montrer les interactions d'objets dans le cadre d'un scénario d'un Diagramme des cas d'utilisation. Pour le but étant de décrire comment se déroulent les actions entre les acteurs ou objets.



**Figure 16** Diagramme sequence pour l'authentification



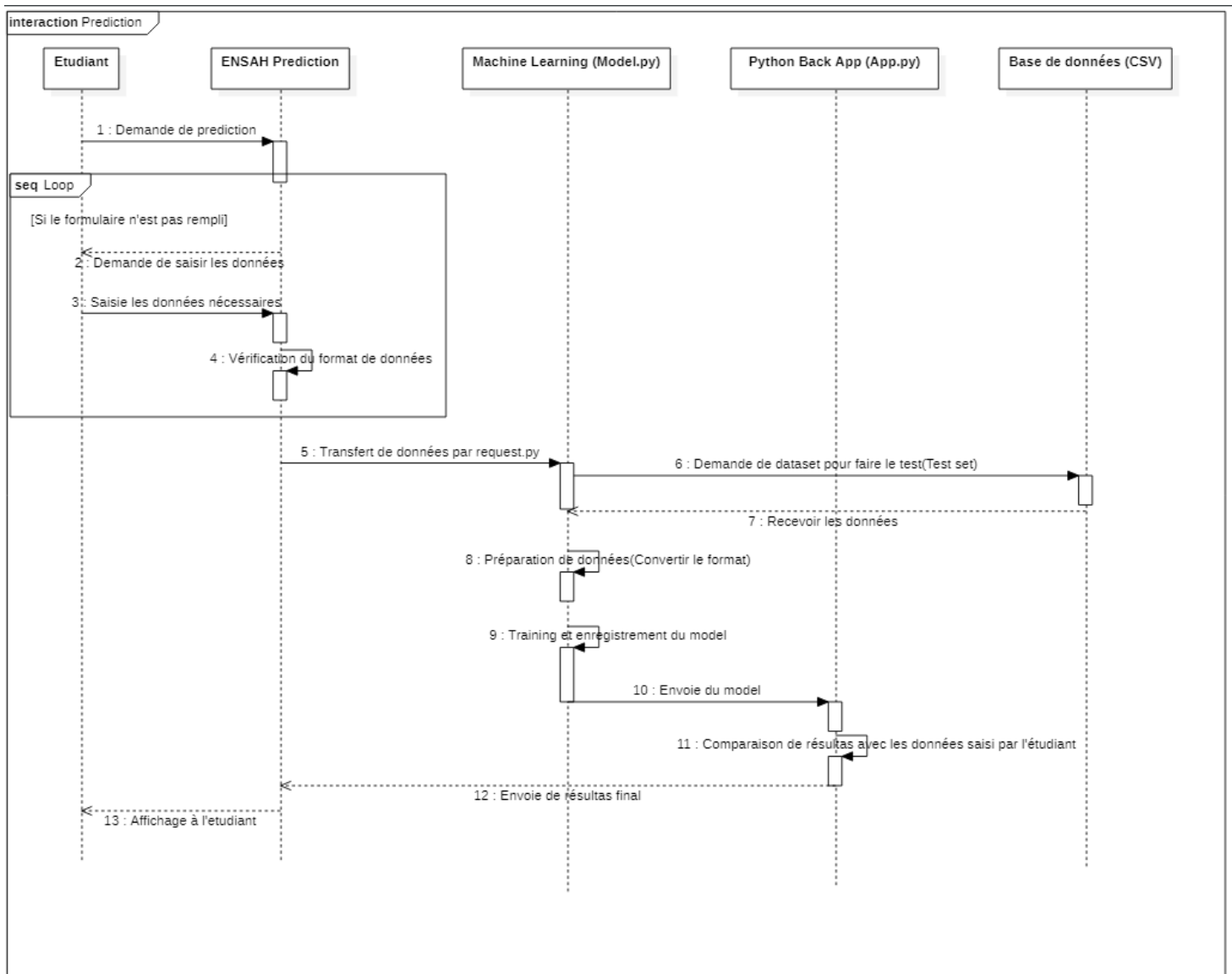


Figure 17 Diagramme de séquence pour la demande de prédiction

## 3.2 Conclusion

Cette étape a été nécessaire pour la structuration de la couche métier. Ainsi nous avons raffiné les diagrammes de classe d'analyse, défini les modules et les composants.

L'étape suivante consiste à l'étude technique du projet afin d'avoir une idée claire sur l'architecture globale.

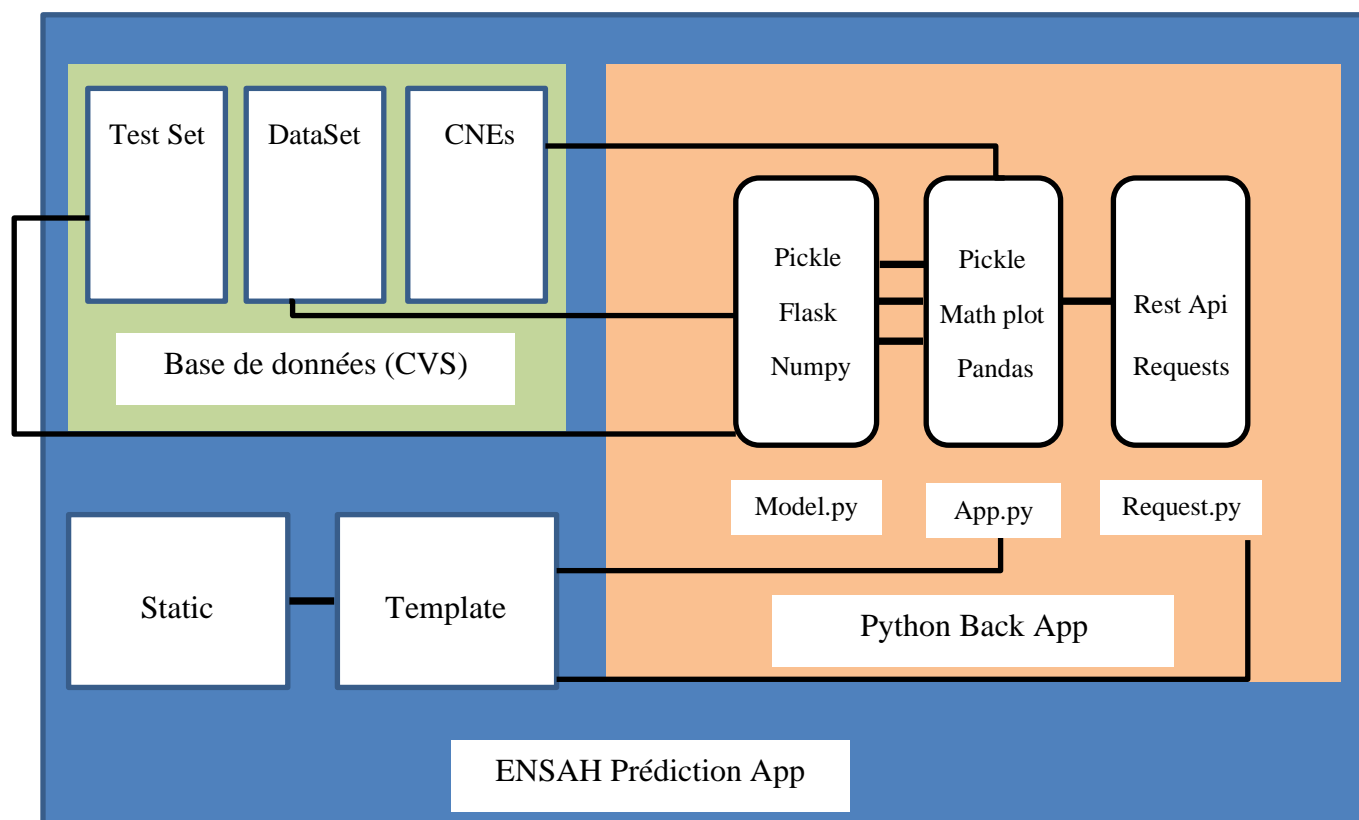
## **Etude Technique**

L'étude technique constitue l'une des étapes de la conception et de l'analyse de faisabilité d'un projet. Elle représente un préalable dans le sens où elle jette les premières bases du projet, c'est la phase d'adaptation de la conception à l'architecture technique retenue, tout en décrivant et documentant le fonctionnement de chaque unité logiciel.

Les étapes de la conception générale des différents modules sont considérées comme les étapes les plus critiques du projet. D'une part, elles doivent être élaborées avec attention et précision, vu que l'architecture à proposer doit rester conforme aux spécifications fonctionnelles. D'autre part elle constitue la base de la réalisation qui concrétise les efforts fournis tout au long du projet.

## 4.1 Architecture technique du projet

L'architecture globale de notre application se résume par ce schéma :



*Figure 18 shema technique du projet*

Comme déjà présenté, le projet est sous forme de deux applications et une base de données interagit entre eux afin de répondre aux requêtes de l'étudiant.

L'application ENSAH Prédiction App reçoit les demandes des étudiants depuis la plateforme et elle transmet les données nécessaires remplis sur le formulaire à Python Back App.

Le service model de l'application Python Back App fait appel à la base de données, il traite les données reçu et crée un modèle et il le transmette au service App pour comparer les résultats avec les données saisi par l'étudiant.

L'application ENSAH Prédiction App reçoit les résultats et affiche le statut final de la prédiction à l'étudiant.

## **4.2 Choix logiciels**

Le développement d'applications web et de sites Internet présente des points communs. D'où l'utilisation des Framework qui vient avec un ensemble cohérent de composants logiciels structurels, qui sert à créer les fondations ainsi que les grandes lignes d'une partie d'un logiciel (architecture). Notre projet se base totalement sur des technologies open-source.

### **4.2.1 Langage Python**

Python est un langage de programmation interprété, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes et d'un système de gestion d'exceptions ; il est ainsi similaire à Perl, Ruby, Scheme, Smalltalk et Tcl.

À la fin des années 1980, le programmeur Guido van Rossum, participe au développement du langage de programmation ABC au Centrum voor Wiskunde en Informatica (CWI) d'Amsterdam, aux Pays-Bas. Il travaillait alors dans l'équipe du système d'exploitation Amoeba dont les appels systèmes étaient difficilement interfaçables avec le Bourne shell utilisé comme interface utilisateur. Il estime alors qu'un langage de script inspiré d'ABC pourrait être intéressant comme interpréteur de commandes pour Amoeba.

En 1989, profitant d'une semaine de vacances durant les fêtes de Noël, il utilise son ordinateur personnel pour écrire la première version du langage, et c'est ici où python commence. Fan de la série télévisée Monty Python's Flying Circus, il décide de baptiser ce projet Python.

Le langage Python est placé sous une licence libre proche de la licence BSD et fonctionne sur la plupart des plates-formes informatiques, des smartphones aux ordinateurs centraux, de Windows à Unix avec notamment GNU/Linux en passant par macOS, ou encore Android, iOS, et peut aussi être traduit en Java ou .NET. Il est conçu pour optimiser la productivité des programmeurs en offrant des outils de haut niveau et une syntaxe simple à utiliser.

Il est également apprécié par certains pédagogues qui y trouvent un langage où la syntaxe, clairement séparée des mécanismes de bas niveau, permet une initiation aisée aux concepts de base de la programmation.

Python est un langage qui peut s'utiliser dans de nombreux contextes et s'adapter à tout type d'utilisation grâce à des bibliothèques spécialisées. Il est cependant particulièrement utilisé comme langage de script pour automatiser des tâches simples mais fastidieuses, comme un script qui récupérerait la météo sur Internet ou qui s'intégrerait dans un logiciel de conception assistée

par ordinateur afin d'automatiser certains enchaînements d'actions répétitives (voir la section Adoption). On l'utilise également comme langage de développement de prototype lorsqu'on a besoin d'une application fonctionnelle avant de l'optimiser avec un langage de plus bas niveau. Il est particulièrement répandu dans le monde scientifique, et possède de nombreuses bibliothèques optimisées destinées au calcul numérique.

### 4.2.2 Framework Flask

Flask est un framework open-source de développement web en Python. Son but principal est d'être léger, afin de garder la souplesse de la programmation Python, associé à un système de templates. Il est distribué sous licence BSD.

Flask a été créé initialement par Armin Ronacher comme étant un poisson d'avril. Le souhait de Ronacher était de réaliser un framework web contenu dans un seul fichier Python mais pouvant maintenir des applications très demandées.

En 2018, Flask était élu "Framework web le plus populaire" par le Python Developers Survey. En janvier 2020, il cumulait plus de 49000 étoiles sur Github, plus que n'importe quel autre framework de développement web Python.

### 4.2.3 Numpy

NumPy est une extension du langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.

Plus précisément, cette bibliothèque logicielle libre et open source fournit de multiples fonctions permettant notamment de créer directement un tableau depuis un fichier ou au contraire de sauvegarder un tableau dans un fichier, et manipuler des vecteurs, matrices et polynômes.

NumPy est la base de SciPy, regroupement de bibliothèques Python autour du calcul scientifique.

### 4.2.4 Scikit-learn

Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique. Elle est développée par de nombreux contributeurs<sup>2</sup> notamment dans le monde académique par des instituts français d'enseignement supérieur et de recherche comme Inria<sup>3</sup>. Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des

algorithmes de classification, et les machines à vecteurs de support. Elle est conçue pour s'harmoniser avec d'autres bibliothèques libres Python, notamment NumPy et SciPy.

#### **4.2.5 Pandas**

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles. Pandas est un logiciel libre sous licence BSD.

Les principales structures de données sont les séries (pour stocker des données selon une dimension - grandeur en fonction d'un index), les DataFrames (pour stocker des données selon 2 dimensions - lignes et colonnes), les Panels (pour représenter des données selon 3 dimensions, les Panels4D ou les DataFrames avec des index hiérarchiques aussi nommés MultiIndex (pour représenter des données selon plus de 3 dimensions - hypercube)

#### **4.2.6 Matplotlib**

Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques<sup>5</sup>. Elle peut être combinée avec les bibliothèques python de calcul scientifique NumPy et SciPy<sup>6</sup>. Matplotlib est distribuée librement et gratuitement sous une licence de style BSD<sup>4</sup>. Sa version stable actuelle (la 2.0.1 en 2017) est compatible avec la version 3 de Python.

Plusieurs points rendent cette bibliothèque intéressante :

- Export possible en de nombreux formats matriciels (PNG, JPEG...) et vectoriels (PDF, SVG...)
- Documentation en ligne en quantité, nombreux exemples disponibles sur internet.
- Forte communauté très active.
- Interface pylab : reproduit fidèlement la syntaxe MATLAB.
- Bibliothèque haut niveau : idéale pour le calcul interactif.

#### **4.2.7 Pickle**

Pickle est un module de python qui permet de sauvegarder une ou plusieurs variables dans un fichier et de récupérer leurs valeurs ultérieurement. Les variables peuvent être de type quelconque.

En clair, si pour une raison quelconque, dans un script **Python**, vous avez besoin de sauvegarder, temporairement ou même de façon plus pérenne, le contenu d'un objet Python comme une liste, un dictionnaire, un tuple etc etc ... au lieu d'utiliser une base de données ou un simple fichier texte, le module **pickle** est fait pour ça.

Il fonctionne comme le module JSON mais n'est pas limité à un seul format d'objet.

### 4.3 Conclusion

Durant ce chapitre, nous avons traité l'étude technique du projet, au cours de laquelle nous avons détaillé les caractéristiques techniques de la solution escomptée. Ensuite, nous avons vu l'architecture technique de l'outil logiciel qui sera utilisé dans la réalisation et mise en œuvre.

# [5<sup>ème</sup>

## CHAPITRE

# Réalisation Et Mise En Œuvre

Dans ce chapitre, nous présenterons les étapes de la réalisation et la mise en œuvre de notre projet.

## 5.1 Outils de développement

Le choix de l'environnement technique du travail est une étape primordiale avant d'entamer la phase de l'implémentation. Cette partie décrit l'ensemble des outils technologiques utilisés lors de développement du système.

### 5.1.1 Training Application Python Back App

Le développement de l'application a exigé l'utilisation de plusieurs outils qui s'est interagissaient entre eux afin de rendre les résultats désirés.



### 5.1.1.1 IDE Python

Pour la programmation en Python, il y a plusieurs fonctionnalités qui facilitent le codage : L'enregistrement et le rechargement des fichiers de code, observer le comportement de votre programme au moment de l'exécution et de détecter les problèmes. Dans notre projet on a travaillé avec Spyder.

Spyder (nommé Pydee dans ses premières versions) est un environnement de développement intégré (IDE) créé par Pierre Raybaut en 2008. Il est multi-plateforme et est sous licence non copyleft (Un logiciel libre non copyleft est diffusé par son auteur avec la permission de le redistribuer et de le modifier). Depuis 2012, c'est la communauté Python scientifique qui maintient Spyder grâce à leur contribution. Il possède les fonctionnalités de base (Coloration syntaxique, autocomplétion...) et intègre de nombreuses bibliothèques d'usage scientifique tels que Matplotlib, NumPy, SciPy, IPython... Il est open source et gratuit et s'installe très facilement grâce au gestionnaire de paquet Python pip

En comparaison avec d'autres IDE pour le développement scientifique, Spyder a un ensemble unique de fonctionnalités - multiplateforme, open-source, écrit en Python et disponible sous une licence non-copyleft. Spyder est extensible avec des plugins, comprend le support d'outils interactifs pour l'inspection des données et incorpore des instruments d'assurance de la qualité et d'introspection spécifiques au code Python, tels que Pyflakes, Pylint et Rope.

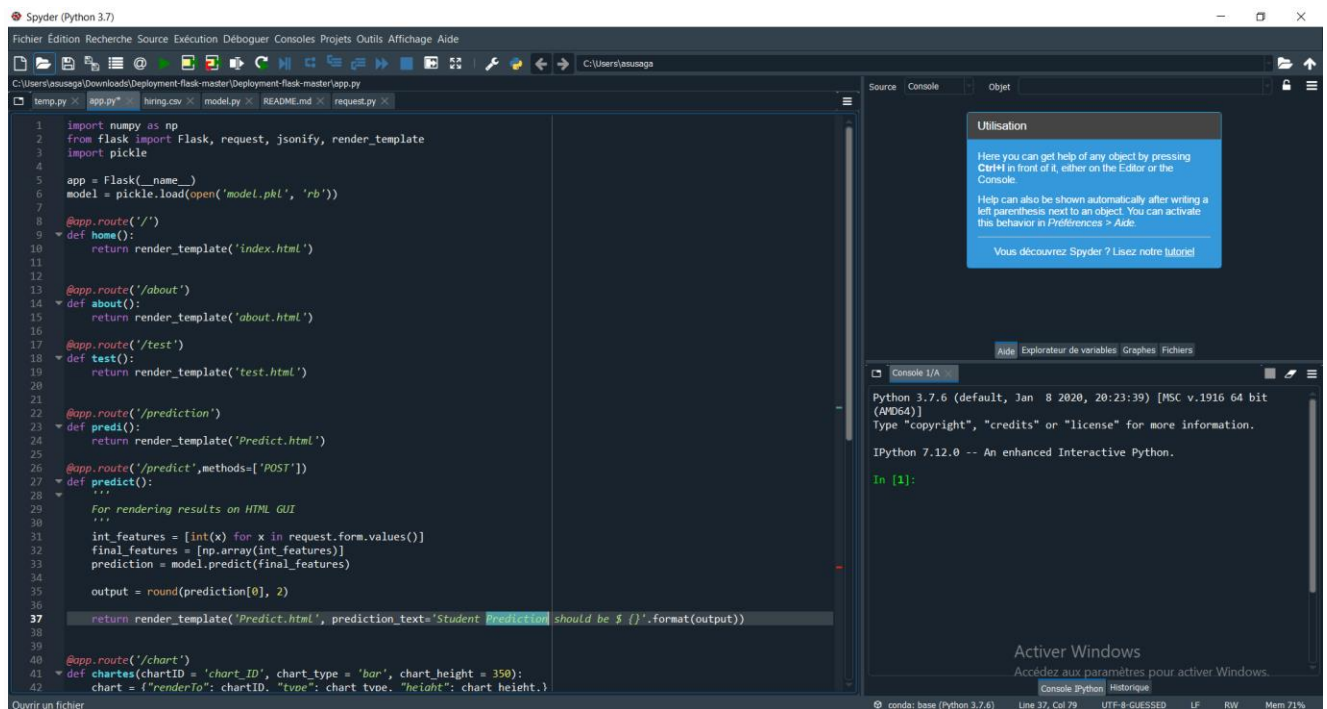
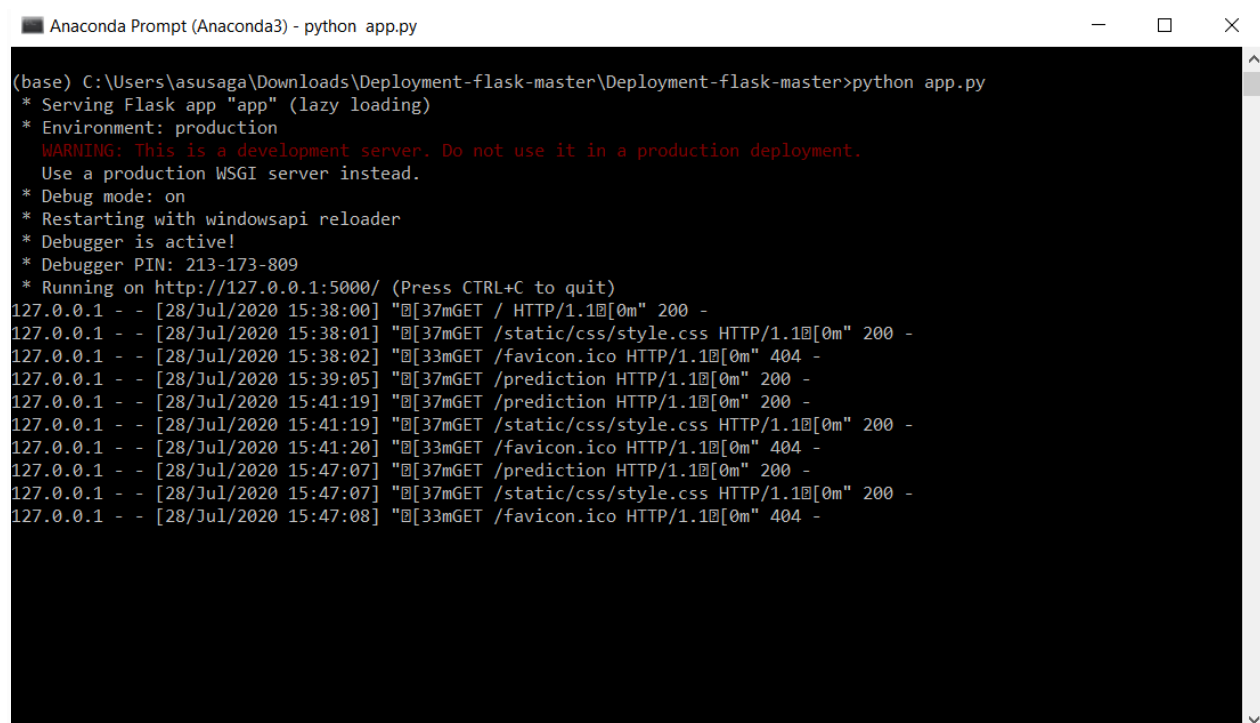


Figure 19 l'editeur spyder

### 5.1.1.2 Python Distribution : Anaconda

Anaconda est une distribution libre et open source<sup>3</sup> des langages de programmation Python et R appliqué au développement d'applications dédiées à la science des données et à l'apprentissage automatique (traitement de données à grande échelle, analyse prédictive, calcul scientifique), qui vise à simplifier la gestion des paquets et de déploiement<sup>4</sup>. Les versions de paquetages sont gérées par le système de gestion de paquets conda<sup>5</sup>. La distribution Anaconda est utilisée par plus de 6 millions d'utilisateurs et comprend plus de 250 paquets populaires en science des données adaptés pour Windows, Linux et MacOS



```
Anaconda Prompt (Anaconda3) - python app.py
(base) C:\Users\asusaga\Downloads\Deployment-flask-master\Deployment-flask-master>python app.py
* Serving Flask app "app" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with windowsapi reloader
* Debugger is active!
* Debugger PIN: 213-173-809
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [28/Jul/2020 15:38:00] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:38:01] "[37mGET /static/css/style.css HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:38:02] "[33mGET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [28/Jul/2020 15:39:05] "[37mGET /prediction HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:41:19] "[37mGET /prediction HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:41:19] "[37mGET /static/css/style.css HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:41:20] "[33mGET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [28/Jul/2020 15:47:07] "[37mGET /prediction HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:47:07] "[37mGET /static/css/style.css HTTP/1.1" 200 -
127.0.0.1 - - [28/Jul/2020 15:47:08] "[33mGET /favicon.ico HTTP/1.1" 404 -
```

*Figure 20 Anaconda Prompt*

### 5.1.1.3 Base de données : Microsoft Excel

**Microsoft Excel** est un logiciel tableur de la suite bureautique Microsoft Office développé et distribué par l'éditeur Microsoft. La version la plus récente est Excel 2019.

Il est destiné à fonctionner sur les plates-formes Microsoft Windows, Mac OS X, Android ou Linux (moyennant l'utilisation de Wine). Le logiciel Excel intègre des fonctions de calcul numérique, de représentation graphique, d'analyse de données (notamment de tableau croisé

dynamique) et de programmation, laquelle utilise les macros écrites dans le langage VBA (Visual Basic for Applications) qui est commun aux autres logiciels de Microsoft Office.

Depuis sa création au début des années 1980 mais surtout à partir de sa version 5 (en 1993), Excel a connu un grand succès tant auprès du public que des entreprises prenant une position très majoritaire face aux logiciels concurrents, tel Lotus 1-2-3. Les principaux formats de fichiers natifs portent l'extension xls (1995→2003) etxlsx (2007→2016). Chaque fichier correspond à un classeur, lequel contient des feuilles de calculs organisées.

Chaque feuille correspond à un tableau de lignes et de colonnes pouvant contenir des valeurs (numériques ou non) ainsi que des formules permettant les calculs. Suivant les versions, les classeurs peuvent aussi inclure les éléments de programmation que sont les macros. Depuis avril 2014, l'application Excel 2013 est disponible sur iPad, iPhone, Windows Phone et sur OS Android depuis 2015. Cela entre dans la logique "Office mobile" de Microsoft de rendre ses logiciels disponibles sur le plus grand nombre de supports (ordinateurs, tablettes, smartphones).

La dernière version en date est la version 2019, elle fait partie de la suite Microsoft Office 2019 et est incluse dans l'abonnement à Microsoft 365. Excel 2019 regorge de nouvelles fonctionnalités tels que,des nouvelles fonctions ( MAX.SI.ENS...),des nouveaux graphiques, des éléments graphiques améliorés, la possibilité également de publier via Power BI.

#### **5.1.1.4 Comma-separated values : CSV**

*Comma-separated values*, connu sous le sigle CSV, est un format texte ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules. Ce format n'a jamais vraiment fait l'objet d'une spécification formelle. Toutefois, la RFC 41801 décrit la forme la plus courante et établit son type MIME « text/csv », enregistré auprès de l'IANA.

Un fichier CSV est un fichier texte, par opposition aux formats dits « binaires ». Chaque ligne du texte correspond à une ligne du tableau et les virgules correspondent aux séparations entre les colonnes. Les portions de texte séparées par une virgule correspondent ainsi aux contenus des cellules du tableau. Une ligne est une suite ordonnée de caractères terminée par un caractère de fin de ligne (line break – LF ou CRLF), la dernière ligne pouvant en être exemptée.

	B	C	D	E	F	G	H	I	J	K
1	Sexe	vous Age ?	Je suis	Filière	ton meilleur endroit d'étude	Vous travaillez en équipe	À quelle heure vous dormez	Combien de fois par semaine est-ce que vous participez	Qu'avez-vous pu	
2	Homme		21 Interne	GI	Café	Non	12:00:00 AM 2 fois	Oui	correct	
3	Homme		18 Interne	CP1	La Bibliothèque	Non	02:00:00	5 Oui	trop lent	
4	Homme		21 Interne	GI	Café	Oui	1:00:00 AM 3 fois	Oui	trop lent	
5	Homme		18 Externe	CP2	Café	Oui	02:00:00	5 Oui	correct	
6	Homme		21 Interne	GC	Chez moi	Oui	01:00:00 3 fois	Non	trop lent	
7	Homme		20 Externe	CP2	Café	Oui	2:00:00 AM 3 fois	Oui	correct	
8	Homme		22 Externe	GI	Café	Non	03:00:00 une fois	Non	trop lent	
9	Homme		20 Interne	CP2	Chez moi	Oui	00:00:00 une fois	Non	trop intense	
10	Homme		20 Interne	GI	Café	Non	02:00:00 jamais	Non	correct	
11	Homme		23 Interne	GI	Chez moi	Oui	01:30:00 une fois	Oui	correct	
12	Homme		23 Externe	GEER	Chez moi	Non	2:00:00 AM jamais	Non	trop lent	
13	Femme		22 Interne	GI	Café	Oui	02:00:00 jamais	Non	trop intense	
14	Femme		22 Interne	GI	La Bibliothèque	Non	12:00:00 3 fois	Non	trop intense	
15	Homme		21 Externe	GI	Chez moi	Oui	12:00:00 AM 3 fois	Non	trop lent	
16	Homme		21 Externe	GC	Chez moi	Oui	03:00:00 3 fois	Oui	correct	
17	Homme		22 Interne	GI	Chez moi	Oui	03:00:00 3 ou plus	Oui	trop lent	
18	Homme		22 Externe	GI	Café	Oui	04:00:00 jamais	Non	trop lent	
19	Femme		21 Interne	GI	La Bibliothèque	Non	01:00:00 2 fois	Oui	trop intense	
20	Homme		22 Externe	GEE	Café	Oui	01:00:00 une fois	Oui	trop lent	
21	Homme		23 Externe	GI	Café	Non	03:00:00 jamais	Non	trop intense	
22	Homme		21 Interne	GI	Café	Non	02:00:00 3 fois	Oui	trop lent	
23	Homme		19 Interne	CP2	La Bibliothèque	Non	11:00:00 une fois	Oui	trop lent	
24	Homme		21 Interne	GI	Chez moi	Non	12:00:00 AM 4 fois	Oui	trop intense	
25	Homme		23 Interne	GI	Café	Non	11:00:00 AM	7	trop lent	
26	Homme		19 Interne	CP2	Chez moi	Non	11:30:00 PM 3 fois	Oui	trop lent	
27	Homme		20 Interne	GI	Café	Non	4:00:00 AM jamais	Oui	trop lent	
28	Homme		22 Interne	GI	Chez moi	Non	1:00:00 AM Occasionnellement	Non	trop intense	
29	Homme		23 Externe	GI	Café	Non	1:00:00 AM 2 fois	Oui	correct	

Figure 21 Feuille Excel

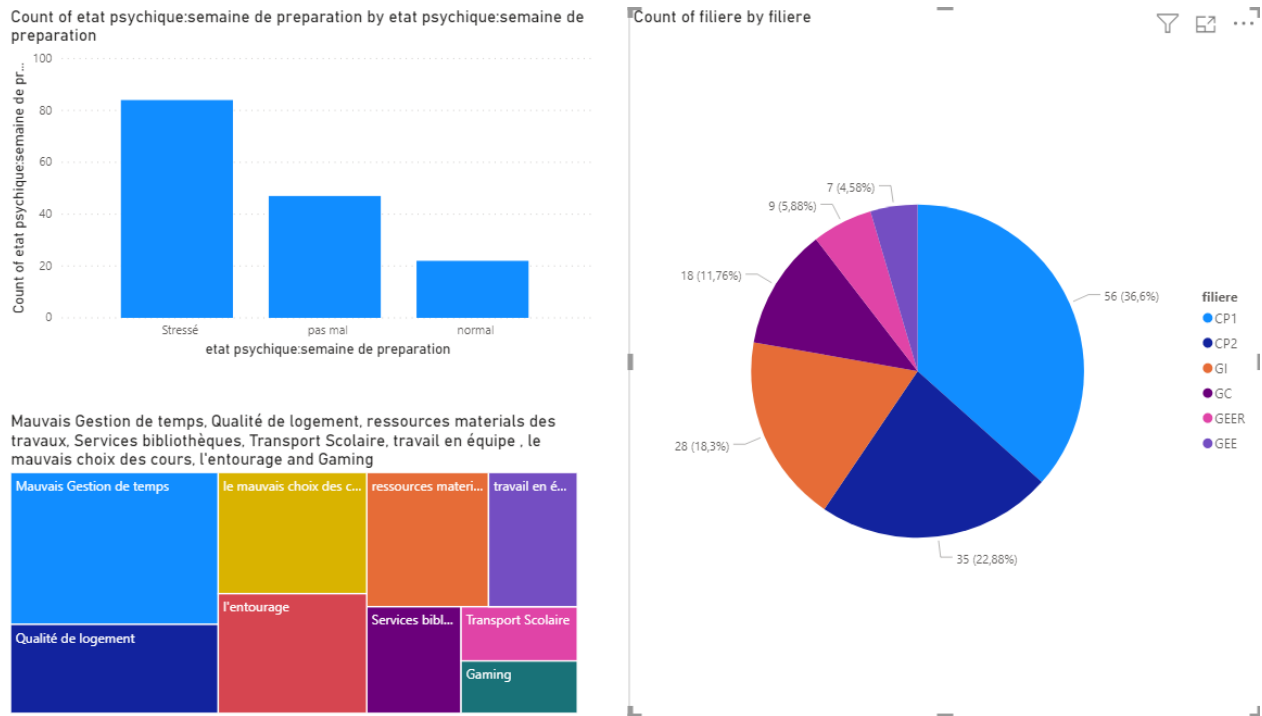
### 5.1.1.5 Power BI

**Power BI** est un service d'analyse commerciale de Microsoft. Il vise à fournir des visualisations interactives et des capacités de veille stratégique avec une interface suffisamment simple pour que les utilisateurs finaux puissent créer leurs propres rapports et tableaux de bord.

Power BI fournit des services de BI (Business Intelligence) basés sur le cloud, appelés «services Power BI», ainsi qu'une interface basée sur le bureau, appelée «Power BI Desktop». Il offre des capacités d'entrepôt de données, notamment la préparation des données, la découverte de données et des tableaux de bord interactifs. En mars 2016, Microsoft a publié un service supplémentaire appelé Power BI Embedded sur sa plateforme cloud Azure. L'un des principaux facteurs de différenciation du produit est la possibilité de charger des visualisations personnalisées.

Cette application a été conçue à l'origine par Thierry D'Hers et Amir Netz de l'équipe SQL Server Reporting Services chez Microsoft. Il a été conçu à l'origine par Ron George à l'été 2010 et nommé Project Crescent. Le projet Crescent était initialement disponible pour téléchargement

public le 11 juillet 2011 avec le nom de code de SQL Server Denali. Plus tard renommé en Power BI, il a ensuite été dévoilé par Microsoft en septembre 2013 sous le nom de Power BI pour Office 365. La première version de Power BI était basée sur les compléments basés sur Microsoft Excel: Power Query, Power Pivot et Power View. Avec le temps, Microsoft a également ajouté de nombreuses fonctionnalités supplémentaires telles que les questions et réponses, la connectivité des données au niveau de l'entreprise et les options de sécurité via les passerelles Power BI. Power BI a été lancé pour la première fois au grand public le 24 juillet 2015.



**Figure 22** Creation du dashboard sur Power Bi

## 5.1.2 ENSAH Prédiction Front

L'application ENSAH Prédiction App est la partie front de notre projet PFA. Cette partie n'a pas nécessité beaucoup d'outils de développement.

Le développement de l'application a été fait à l'aide de l'éditeur texte Sublime Text.

### 5.1.2.1 Sublime Text

Sublime Text est un éditeur de texte générique codé en C++ et Python, disponible sur Windows, Mac et Linux. Le logiciel a été conçu tout d'abord comme une extension pour Vim, riche en fonctionnalités.

Depuis la version 2.0, sortie le 26 juin 2012, l'éditeur prend en charge 44 langages de programmation majeurs, tandis que des plugins sont souvent disponibles pour les langages plus rares.

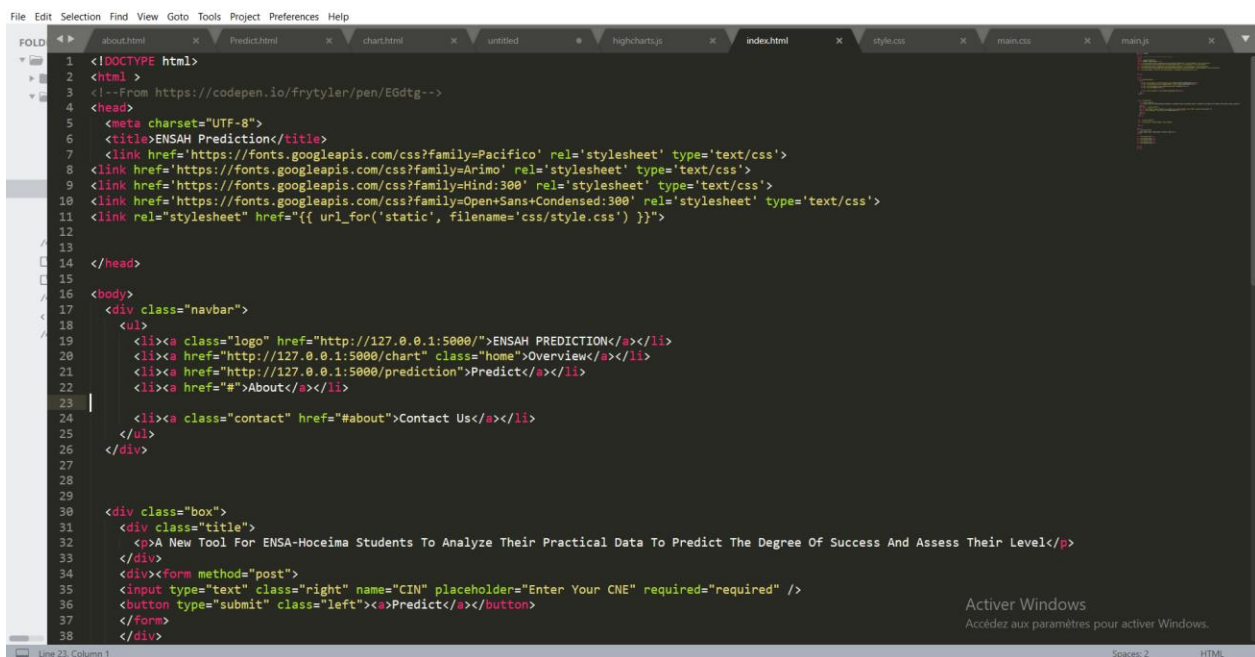


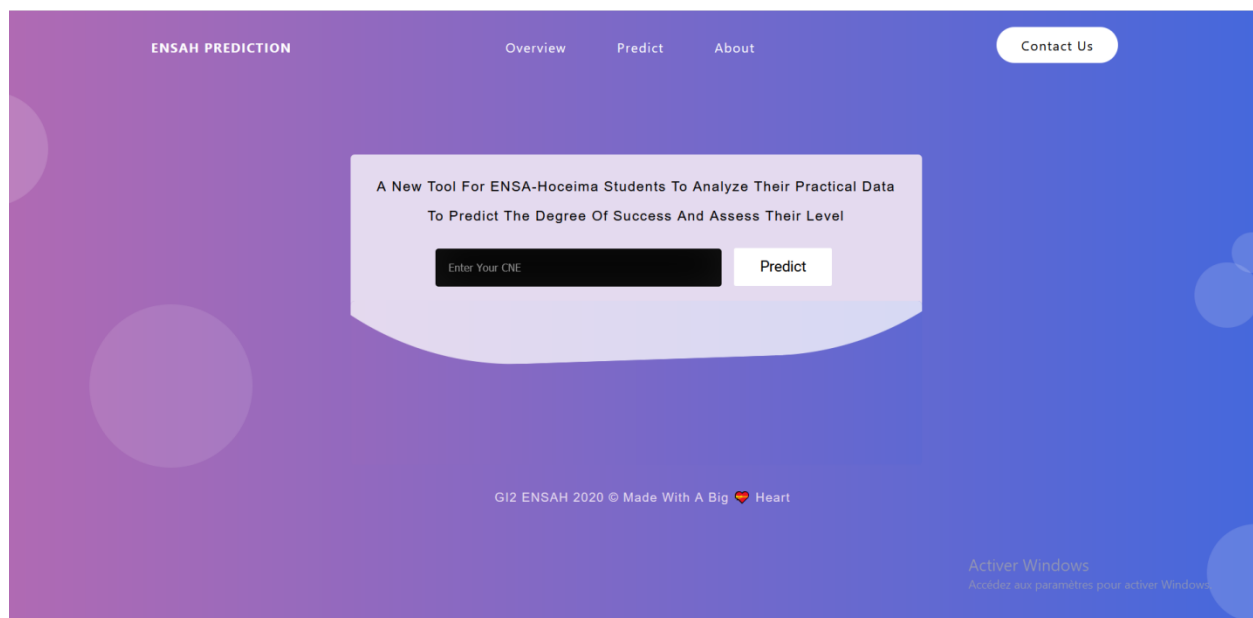
Figure 23 L'éditeur Sublime text

## 5.2 Présentation de la plateforme ENSAH Prédiction.

Nous présenterons dans cette section un aperçu général sur le fonctionnement de la plateforme en utilisant quelques captures d'écran.

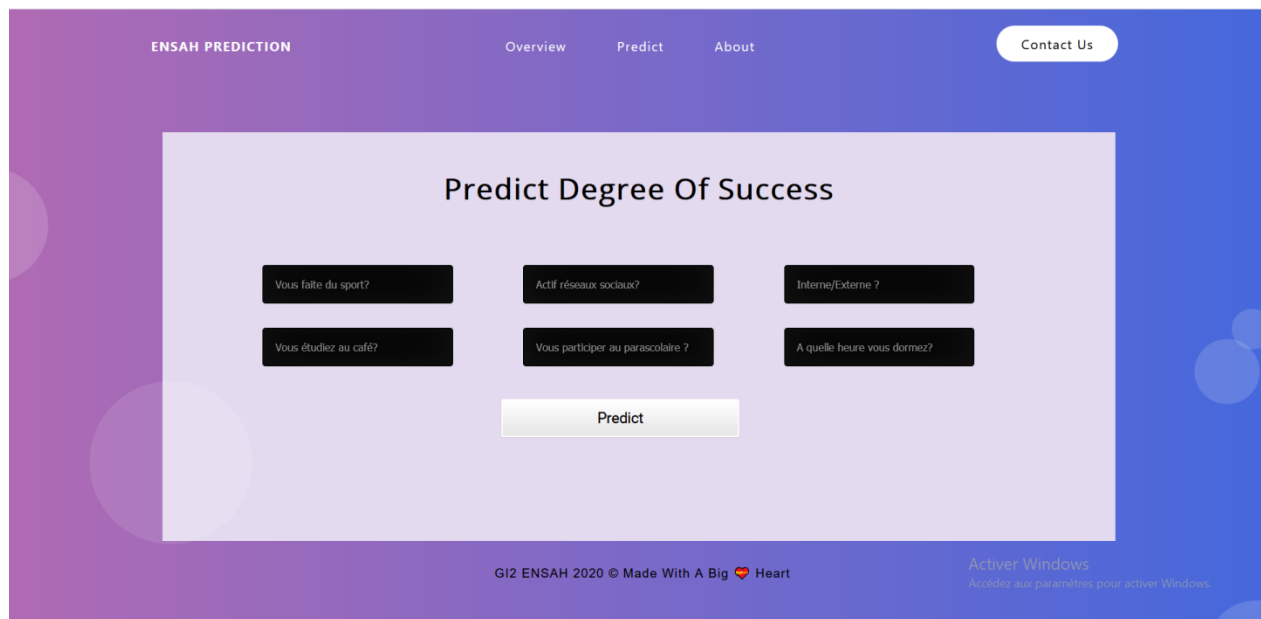
La première page qu'un utilisateur peut visiter est la page d'accueil :

Cette page n'est qu'une page introductive au projet. Elle contient des liens rapides vers les pages majeures de la plateforme et un champ de formulaire pour remplir le CNE de l'étudiant et qui envoie vers la page prédiction.



**Figure 24** Page d'accueil

Pour la demande de prédiction, l'étudiant est demandé d'insérer les données et remplir le formulaire. Ce processus se fait à l'aide de la page Prédiction.



*Figure 25 Page de prédiction*

### 5.3 Conclusion

Notre plateforme offre un service excellent et innovant pour les élèves de l'ENSAH, tout le projet a été développée en utilisant les technologies les plus récentes et les plus réputées dans le monde de l'IT au niveaux de toutes les phases de développement.



# Annexe

Annexe 1 : Questionnaire distribué dans la première phase d'enquête

## QUESTIONNAIRE SUR LES CONDITIONS DE VIE ET DE TRAVAIL DES ETUDIANTS

Une seule réponse possible.

☐ Homme

☐ Femme

2. votre Age ? \*

\_\_\_\_\_

3. Je suis ..

Une seule réponse possible.

☐ Interne

☐ Externe

4. Filiere

Une seule réponse possible.

☐ CP1

☐ CP2

☐ GC

☐ GEER

☐ GEE

☐ GI

[https://docs.google.com/forms/d/13TFMfdcp2MU\\_4a7OnVB8CbBqQlUX7uJ/KGGQc\\_IBiA/edit](https://docs.google.com/forms/d/13TFMfdcp2MU_4a7OnVB8CbBqQlUX7uJ/KGGQc_IBiA/edit) 1/5

---

7/30/2020 Vos coordonnées

5. ton meilleur endroit d'étudier . ? \*

Une seule réponse possible.

☐ Chez moi

☐ La Bibliothèque

☐ Café

6. Vous travaillez en équipe ! \*

*Une seule réponse possible.*

- ☐ Oui  
☐ Non

7. A quelle heure vous dormez ? \*

*Exemple : 8 h 30*

8. Combien de fois par semaine vous faites de sport ? \*

*Une seule réponse possible.*

- ☐ jamais  
☐ une fois  
☐ 2 fois  
☐ 3 fois  
☐ Occasionnellement

9. est ce que Vous participez aux evenements parascolaires ! \*

*Une seule réponse possible.*

- ☐ Oui  
☐ Non

[https://docs.google.com/forms/d/13TFMfdcpe2MJ\\_4a7OnVB8CbBqQlJX7uKGGQc\\_JBiI/edit](https://docs.google.com/forms/d/13TFMfdcpe2MJ_4a7OnVB8CbBqQlJX7uKGGQc_JBiI/edit)

2/5

7/30/2020

Vos coordonnées

10. Qu'avez-vous pensé du rythme de la formation ? \*

*Une seule réponse possible.*

- ☐ trop lent  
☐ correct  
☐ trop intense

11. comment évaluez-vous les compétences démontrées par les enseignants ? \*

*Une seule réponse possible.*

- ☐ très élevées
- ☐ élevées
- ☐ pas mal
- ☐ faibles
- ☐ très faibles

12. Êtes-vous satisfait de la communication avec les enseignants ? \*

*Une seule réponse possible.*

- ☐ Satisfait
- ☐ Moyennement satisfait
- ☐ Insatisfait

13. Êtes vous satisfait des méthodes pédagogiques proposées ? \*

*Une seule réponse possible.*

- ☐ Satisfait
- ☐ Moyennement satisfait
- ☐ Insatisfait

[https://docs.google.com/forms/d/13TFMldcpe2MJ\\_4a7OnVB8CbBqQIJX7ujKGQc\\_IBiA/edit](https://docs.google.com/forms/d/13TFMldcpe2MJ_4a7OnVB8CbBqQIJX7ujKGQc_IBiA/edit)

3/5

7/30/2020

Vos coordonnées

14. Evaluer votre état psychique pendant la semaine de preparation ? \*

*Une seule réponse possible.*

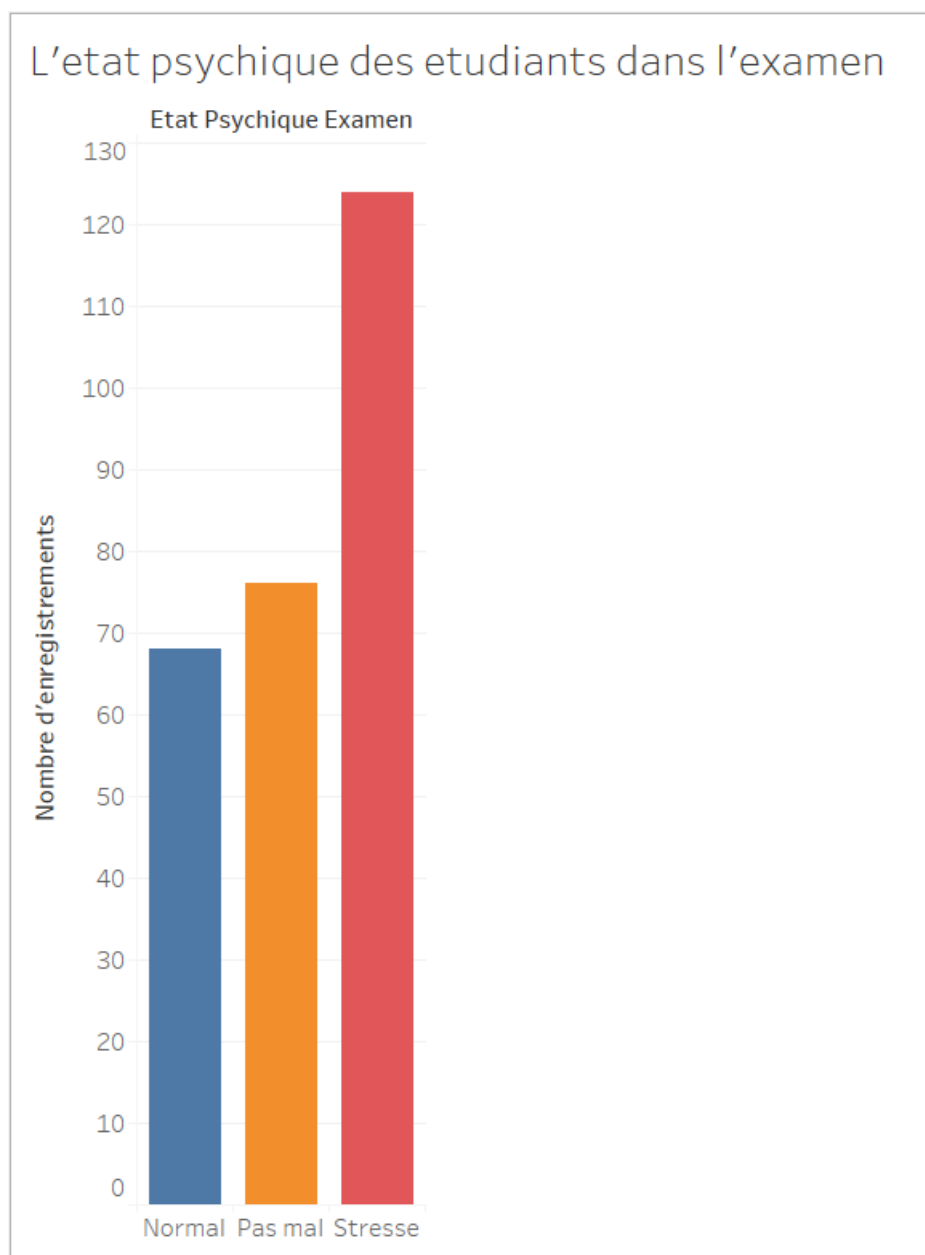
- ☐ normal
- ☐ pas mal
- ☐ Stressé

18. Lesquels , parmi les facteurs suivants , ont influencés negativement l'interet a vos etudes ? \*

*Une seule réponse possible par ligne.*

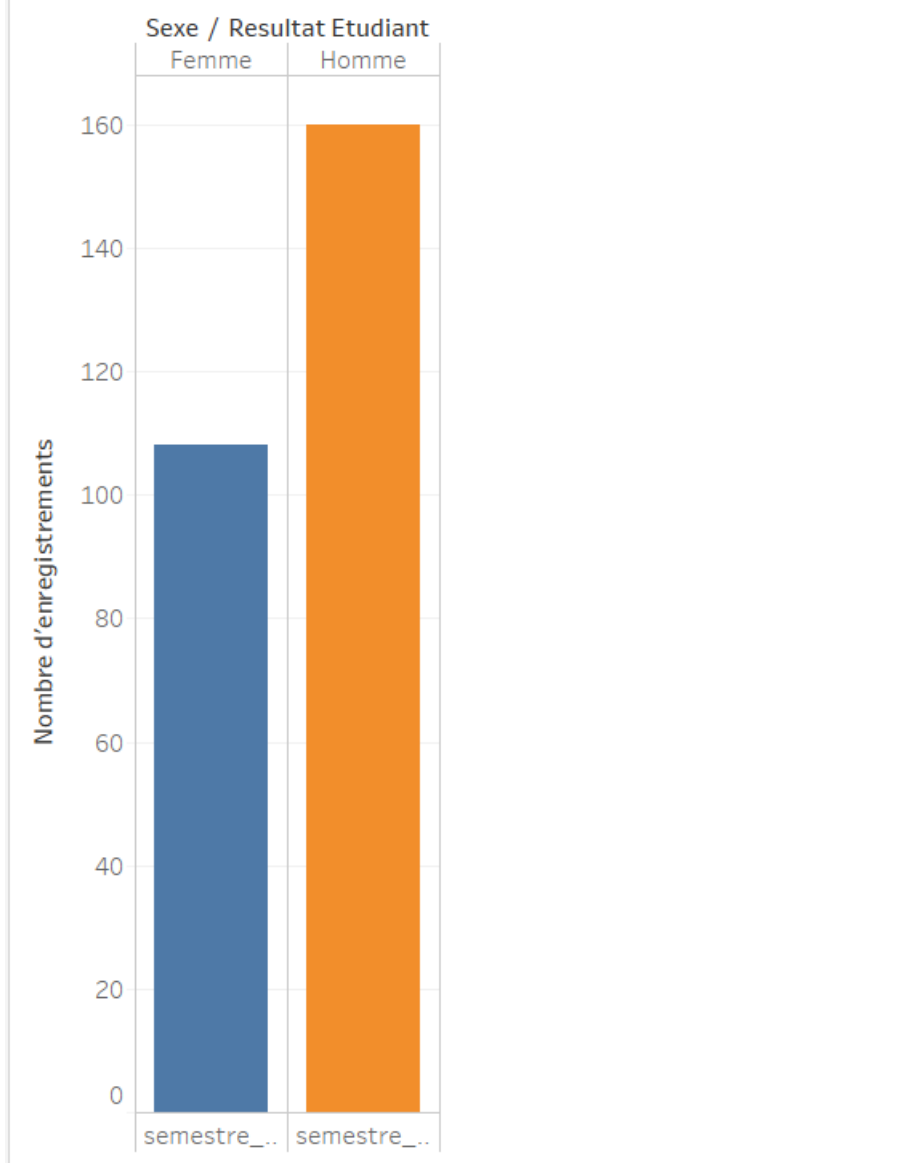
	Oui	Non
Services bibliothèques	<input type="radio"/>	<input type="radio"/>
Transport Scolaire	<input type="radio"/>	<input type="radio"/>
Orientation	<input type="radio"/>	<input type="radio"/>
ressources materials des travaux	<input type="radio"/>	<input type="radio"/>
Mauvais Gestion de temps	<input type="radio"/>	<input type="radio"/>
Qualité de logement	<input type="radio"/>	<input type="radio"/>
l'addiction aux drogues	<input type="radio"/>	<input type="radio"/>
le mauvais choix des cours	<input type="radio"/>	<input type="radio"/>
Les réseaux sociaux	<input type="radio"/>	<input type="radio"/>
Barrières relies au langage	<input type="radio"/>	<input type="radio"/>
l'entourage	<input type="radio"/>	<input type="radio"/>
Gaming	<input type="radio"/>	<input type="radio"/>

## Annexe 2 : Visualisation des données des étudiants



*Figure 26 :L'état psychique des étudiants dans l'examen*

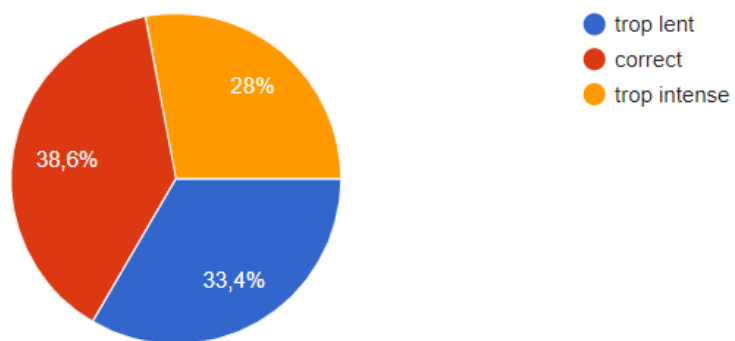
## Semetre non validé selon le sexe



*Figure 27 distrubition de réussite selon le sexe*

Qu'avez-vous pensé du rythme de la formation ?

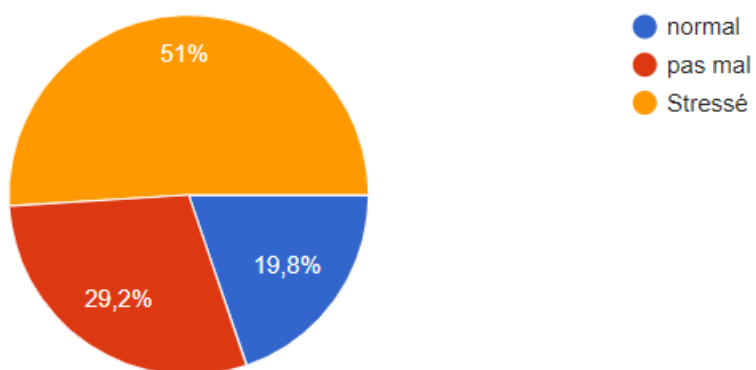
404 réponses



*Figure 28 taux de satisfaction des étudiants par le rythme de formation au sein de l'ensa hoceima*

Evaluer votre etat psychique pendant la semaine de preparation ?

404 réponses



*Figure 29 l'etat psychique des etudiants dans la semaine de préparation*

## Conclusion général

L'objet principal de ce travail de recherche porte sur l'amélioration de la compréhension de la réussite des étudiants de l'école nationale des sciences appliquées Al Hoceima ,en vue d'affronter le désintérêt et l'échecs et d'abandons. La posture adoptée s'inscrit dans une démarche intégrative visant à cerner un ensemble de facteurs influençant la réussite, celle-ci émanant d'une réalité multifactorielle. La prise en compte plus spécifiquement des capacités cognitives des étudiants, évaluées par des tests psychométriques à leur entrée à l'université et la mesure de leur niveau de motivation, apportent à ce travail une dimension originale. A l'issue de cette recherche empirique, des résultats peuvent être dégagés quant à la décomposition des déterminants de la réussite. Les apports et les limites de l'analyse sont également mis en exergue.

L'étudiant intègre l'université avec un « capital » de connaissances, un niveau initial, un parcours scolaire, des capacités cognitives et son « background » personnel. Ces facteurs influencent la réussite fortement ou plus indirectement. La réussite cependant ne peut pas dépendre uniquement de facteurs antérieurs à l'entrée à l'université. L'étudiant est en effet un apprenant actif qui joue un rôle dans son apprentissage et sa réussite. Elle ne peut non plus être liée à des facteurs uniquement cognitifs au cours de l'année universitaire. S'ils sont un élément indispensable à l'apprentissage, ils doivent encore être sollicités par l'étudiant. C'est pourquoi, dans un dernier temps, le rôle de la motivation des étudiants a été appréhendé. Il s'avère, au vu des résultats, c'est la motivation intrinsèque et la motivation qui exercent le plus d'effet, respectivement positif et négatif, sur la réussite universitaire. La partition des étudiants selon qu'ils soient ou non motivés montrent que la motivation permet un gain de 1,8 point sur l'état psychique de l'étudiant pendant l'année universitaire ce qui est un avantage non négligeable.



## Webographie

- <https://www.coursera.org/professional-certificates/ibm-data-science>
- <https://www.python.org/doc/>
- <https://flask.palletsprojects.com/en/1.1.x/>
- <https://fr.wikipedia.org/wiki/Git>
- <https://github.com/andkret/Cookbook/tree/master/sections>