Université de Carthage

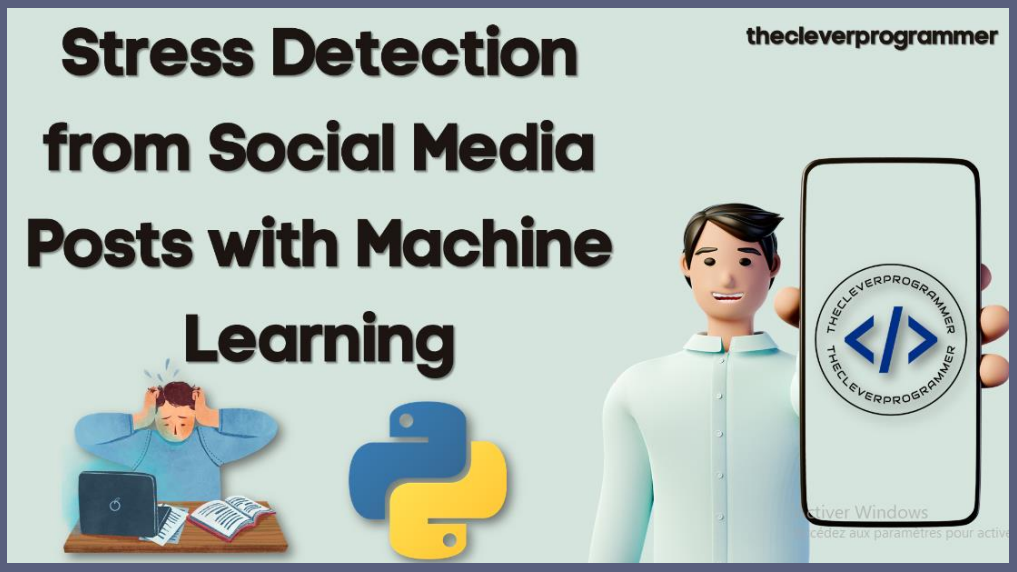
Faculté des Sciences Economiques et de Gestion de Nabeul

Rapport de Mini projet en matière de Machine Learning

Présentées par

Hadil Ben Romdhane
Abir Hazmi

Prédiction de Stress Humain
Analyse des méthodes de classification



Année Universitaire 2022 – 2023

Table des matières

Introduction générale

"Everybody is always under some degree of stress. Even while quietly asleep our heart must continue to beat, our lungs to breathe, and even our brain works in the form of dreams. Stress can be avoided only by dying."

— Hans Selye 'The Nature of Stress', 1985 [283]

La prédiction du stress humain est un domaine de recherche en psychologie et en informatique qui cherche à développer des méthodes pour évaluer et prédire le niveau de stress des individus. Le stress est une réponse normale du corps à des situations de menace ou de pression, mais un niveau de stress élevé peut avoir des effets négatifs sur la santé physique et mentale. La prédiction du stress est donc importante pour aider les individus à gérer leur stress et à prévenir les effets négatifs associés.

Les chercheurs ont utilisé diverses approches pour prédire le stress humain, notamment l'utilisation de capteurs physiologiques tels que les moniteurs de fréquence cardiaque et de respiration pour mesurer les réponses physiologiques au stress, ainsi que l'analyse de données comportementales et d'expression faciale pour déterminer les signes de stress.

Ces données sont souvent analysées à l'aide d'algorithmes d'apprentissage automatique et d'intelligence artificielle pour identifier des modèles qui peuvent être utilisés pour prédire le niveau de stress des individus. La prédiction du stress peut avoir des applications pratiques dans de nombreux domaines, notamment en santé mentale, en médecine du travail et en gestion du stress.

Cependant, la prédiction du stress humain soulève également des questions éthiques, notamment en ce qui concerne la collecte et l'utilisation de données sensibles, la confidentialité et la protection des données, ainsi que la question de savoir si la prédiction du stress peut être utilisée de manière préventive ou coercitive.

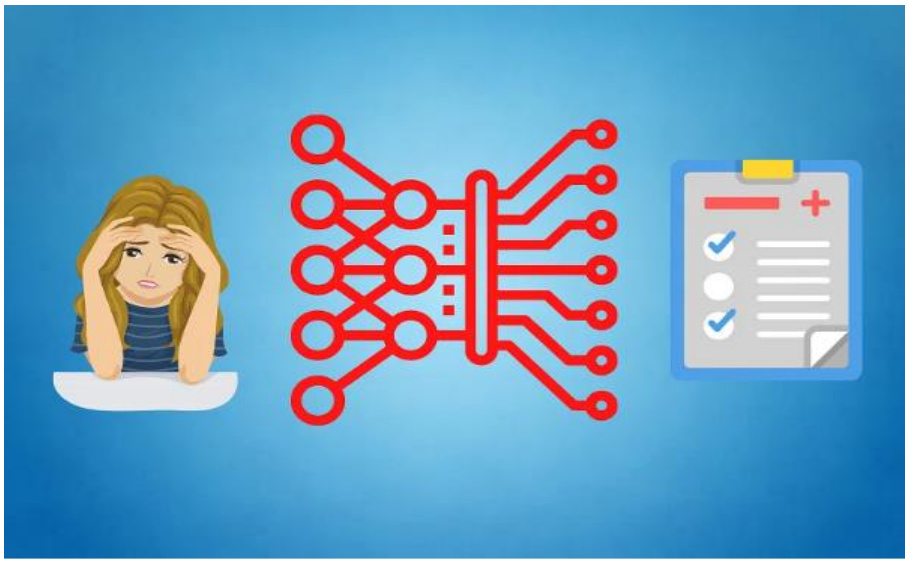
Chapitre I : Présentation générale

I.1. Introduction

Le stress est devenu extrêmement courant avec 33% des personnes souffrant de stress extrême. La technologie peut être utile pour mesurer et gérer le stress.

Le diagnostic des événements stressants et des déclencheurs n'est pas facile en s'appuyant uniquement sur la technologie des capteurs, tels que la conductance cutanée (SC) et la température des doigts (FT), car il y a tellement de mots qui peuvent être utilisés par les gens sur leurs messages qui peuvent montrer si une personne souffre de stress psychologique ou non.

On recherche des ensembles de données que nous pouvons utiliser pour former un modèle d'apprentissage automatique pour la détection du stress.



I.2. Problématique :

Il existe plusieurs problématiques éthiques associées à l'utilisation de techniques de prédiction du stress basées sur la machine Learning.

Tout d'abord, il y a la question de **la collecte et de l'utilisation de données sensibles**. Les données physiologiques et comportementales utilisées pour prédire le stress peuvent être considérées comme des données de santé et donc soumises à des réglementations strictes en matière de protection des données. Il est important de s'assurer que les données sont collectées

de manière éthique et que les individus sont conscients de la manière dont leurs données sont utilisées.

De plus, il est important de veiller à ce que **les algorithmes utilisés pour prédire le stress soient équitables et non biaisés**. Les biais dans les données d'entraînement peuvent entraîner des prédictions incorrectes et des résultats injustes pour certains groupes d'individus. Il est donc essentiel de tester les algorithmes pour détecter les biais et de s'assurer que les résultats sont équitables pour tous les groupes.

Enfin, il est important de se rappeler que la prédiction du stress ne doit pas être utilisée de manière **préventive ou coercitive**. Les individus doivent avoir le droit de refuser de participer à des programmes de prédiction du stress et ne doivent pas être pénalisés pour leur choix. De plus, les résultats de la prédiction du stress doivent être utilisés pour aider les individus à gérer leur stress, plutôt que pour justifier des mesures coercitives ou punitives.

I.3. Solution Proposé :

Les algorithmes de machine Learning peuvent aider à résoudre certaines des problématiques liées à la prédiction de stress.

Tout d'abord, il est important de sélectionner des algorithmes qui sont équitables et non biaisés. Les algorithmes d'apprentissage automatique peuvent être formés de manière à détecter et à corriger les biais dans les données d'entraînement, ce qui peut aider à garantir que les résultats de la prédiction du stress sont équitables pour tous les groupes.

De plus, les algorithmes peuvent être conçus pour protéger la confidentialité des données des individus. Les techniques telles que la cryptographie homomorphe peuvent être utilisées pour permettre aux algorithmes de traiter les données sans avoir accès aux données sensibles réelles, ce qui peut aider à préserver la confidentialité des données tout en permettant des prévisions précises.

Enfin, les algorithmes peuvent être utilisés pour aider les individus à gérer leur stress. Les prévisions de stress peuvent être utilisées pour fournir des recommandations personnalisées aux individus, telles que des techniques de gestion du stress ou des recommandations pour réduire les niveaux de stress. Cela peut aider les individus à mieux comprendre leur stress et à prendre des mesures pour le gérer de manière proactive.

Cependant, il est important de se rappeler que les algorithmes ne sont qu'un outil et que leur efficacité dépendra de la qualité des données d'entrée, de la conception de l'algorithme et de la manière dont il est utilisé. Il est donc essentiel de s'assurer que les algorithmes sont développés et utilisés de manière éthique et responsable.

I.4. Description des données :

Les données dans le fichier **Stress.csv,** contiennent un total de 2838 instances (ou lignes) avec 7 attributs (ou colonnes).Sur ces 7 attributs, 3 sont de nature numérique (entier et réel) et 4 de nature catégorielle (objet).

I.4.1. Fiche signalétique des attributs dans la base de données

Voici une description générale de chaque attribut dans la base de données :

- Subreddit : le nom du sous-reddit auquel appartient le message
- Post_id : l'identifiant unique du message sur Reddit
- Sentence_range : la plage de phrases dans le texte du message qui a été annotée pour l'étiquetage
- Text : le texte de message
- Label : une étiquette ou d'un libellé attribué au message en fonction de son contenu. Il peut s'agir, par exemple, d'une étiquette de sentiment, d'une étiquette de sujet, etc.
- Confidence : la mesure de confiance ou de certitude associée à l'étiquette attribuée au message
- Social_timestamp : la date et de l'heure auxquelles le message a été publié sur Reddit

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2838 entries, 0 to 2837
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   subreddit        2838 non-null   object
 1   post_id          2838 non-null   object
 2   sentence_range   2838 non-null   object
 3   text             2838 non-null   object
 4   label            2838 non-null   int64
 5   confidence       2838 non-null   float64
 6   social_timestamp  2838 non-null   int64
dtypes: float64(1), int64(2), object(4)
memory usage: 155.3+ KB
```

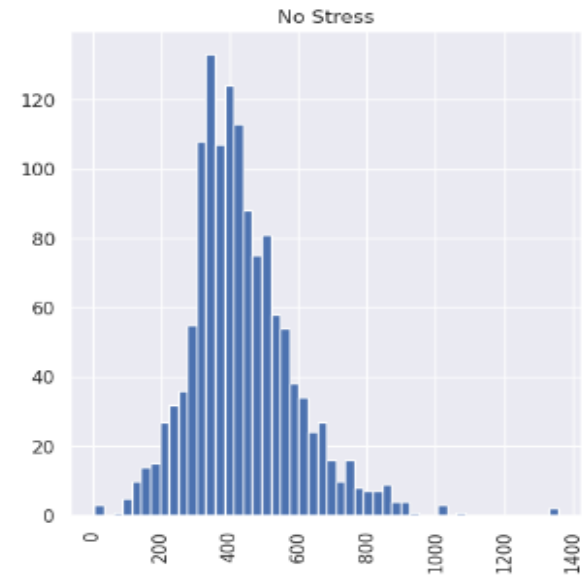
I.4.2. Visualisation des donées

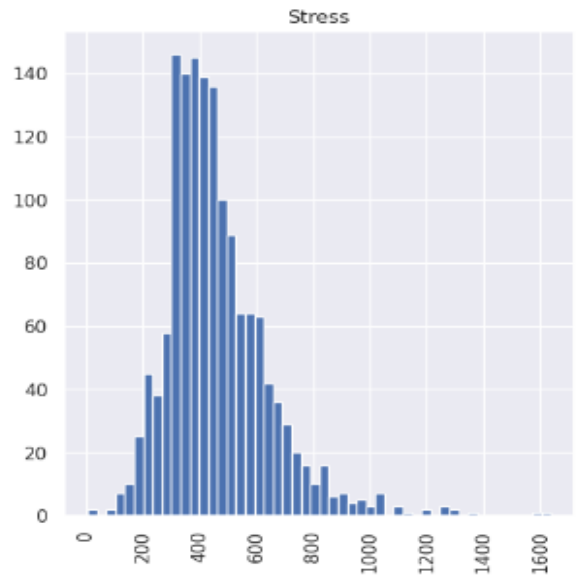
L'ensemble de données contient des données publiées sur des subreddits liés à la santé mentale.

Cet ensemble de données contient divers problèmes de santé mentale partagés par des personnes au sujet de leur vie. Heureusement, cet ensemble est étiqueté comme 0 et 1 ; où 0 indique aucun stress et 1 indique un stress.

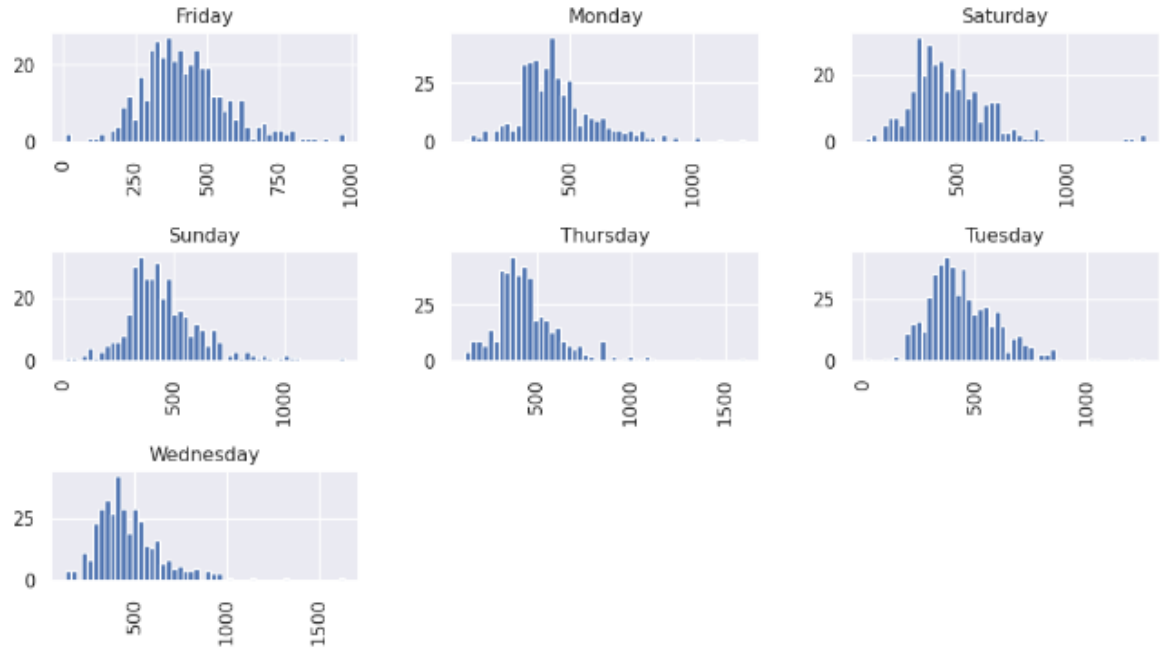
	subreddit	post_id	sentence_range	text	label	confidence	social_timestamp
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	1	0.8	1521614353
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	0	1.0	1527009817
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	1	0.8	1535935605
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing, h...	1	0.6	1516429555
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1	0.8	1539809005
5	relationships	7tx7et	(30, 35)	I think he doesn't want to put in the effort f...	1	1.0	1517274027
6	domesticviolence	7iphly	[25, 30]	It was a big company so luckily I didn't have ...	0	0.8	1512854409
7	anxiety	5m3k80	(5, 10)	It cleared up and I was okay but. On Monday ...	1	0.8	1483582174
8	relationships	7nhy1v	(50, 55)	I actually give an assistant half my emergency...	1	0.6	1514843984
9	assistance	61eiq6	[15, 20]	I just feel like the street life has fucked my...	1	1.0	1490428087

On visualisant les résultats on remarque qu'il n'y a pratiquement aucune différence entre la distribution de stress et la distribution sans stress





Lors de l'ajout d'attribut de date on remarque que Vendredi et mardi pour la longueur du texte sont élevés.



I.6. Organisation de Manuscrit :

Ce manuscrit s'organise en deux sections :

La première section présente un état de l'art axé autour de la notion de stress et de la reconnaissance automatique du stress. Cette section est divisée en trois chapitres : le premier chapitre s'intéresse à la définition du stress d'un point de vue physiologique, psychologique et sociologique ; le deuxième chapitre s'intéresse à la méthodologie de mise en œuvre d'un système de détection automatique du stress et les techniques d'apprentissage automatique en usage pour la reconnaissance du stress.

La deuxième section présente un aperçu des techniques d'apprentissages automatiques utilisés dans la reconnaissance du stress. Cette section est divisé en deux chapitres : le premier chapitre s'intéresse à l'énumération des caractéristiques des techniques ainsi leur avantages et inconvenants ; le deuxième chapitre est conçu pour la comparaison entre les différents méthodes en mentionnant les différences et les simulations.

Chapitre II : Etat de l'art

II.1. Introduction

Dans cette partie, nous présentons un aperçu des travaux existants sur la compréhension et la reconnaissance du stress, de façon pluridisciplinaire et orientée vers la conception d'outils de remédiation. Cette partie se divise en deux chapitres :

Le Chapitre 1 « Le Stress » explicite les notions de stress, de phobie, et de coping (i.e. la gestion du stress) d'un point de vue physiologique et cognitif, et montre l'aspect fortement individuel du stress suivant des facteurs contextuels, socioculturels et personnels.

Le Chapitre 2 « Apprentissage automatique pour la reconnaissance des états affectifs » présente les méthodologies et techniques d'apprentissages automatiques utilisés dans la reconnaissance du stress.

II.2. Le Stress

II.2.1. Définition du stress

Le stress est une réponse physiologique et psychologique du corps à des événements stressants ou à des situations perçues comme menaçantes. Il peut se manifester par des symptômes physiques tels que la tension musculaire, la fatigue, des maux de tête ou des douleurs abdominales, ainsi que par des symptômes émotionnels tels que l'anxiété, la dépression ou l'irritabilité. Le stress peut être à court terme (aigu) ou à long terme (chronique) et peut avoir des effets néfastes sur la santé mentale et physique.

II.2.2. Définition de la phobie

La phobie est un trouble anxieux caractérisé par une peur intense, irrationnelle et persistante d'un objet, d'une situation ou d'une activité spécifique. La personne atteinte de phobie peut avoir des symptômes physiques tels que la sudation, le tremblement, l'accélération du rythme cardiaque ou des nausées lorsqu'elle est confrontée à l'objet de sa phobie. Cette peur peut être si intense que la personne évite toute situation qui pourrait la provoquer, ce qui peut entraîner des limitations importantes dans sa vie quotidienne. Les phobies sont souvent traitées par thérapie cognitivo-comportementale et, dans certains cas, par des médicaments.

II.2.3. Définition de coping

Le coping, (la stratégie d'adaptation), fait référence aux efforts cognitifs et comportementaux que les individus déploient pour faire face à des situations stressantes ou menaçantes. Le coping peut être adaptatif ou mal adaptatif, en fonction de la stratégie adoptée et de son efficacité dans la gestion de la situation stressante. Les stratégies de coping peuvent inclure la recherche de soutien social, la résolution de problèmes, la réévaluation positive, la distraction ou le désengagement. Le choix de la stratégie de coping dépend de facteurs tels que la nature du stressor, les ressources personnelles et sociales disponibles et les préférences individuelles.

II.3. Apprentissage automatique pour la reconnaissance des états affectifs

La prédiction du stress humain est un domaine de recherche relativement récent, mais qui a suscité beaucoup d'intérêt ces dernières années. Les méthodes de prédiction du stress peuvent être classées en plusieurs catégories, en fonction de la source de données utilisée et de la technique d'apprentissage utilisée.

II.3.1. Les méthodes :

Voici une synthèse des principales méthodes utilisées pour la prédiction du stress humain :

a) Méthodes basées sur les signes physiologiques :

Ces méthodes utilisent des données physiologiques pour prédire le stress. Les signes physiologiques tels que la fréquence cardiaque, la respiration, la température de la peau et la conductivité électrodermale peuvent être mesurés à l'aide de capteurs portables ou de dispositifs médicaux. Les données physiologiques peuvent être analysées en utilisant des techniques de traitement du signal et des algorithmes de machine Learning pour prédire le niveau de stress.

b) Méthodes basées sur l'analyse du langage :

Ces méthodes utilisent l'analyse du langage naturel pour prédire le stress. Les données de langage peuvent être collectées à partir de journaux intimes, de journaux de bord électroniques, de conversations en ligne ou d'autres sources de texte. Les données de langage sont analysées en utilisant des techniques de traitement du langage naturel et des algorithmes de machine Learning pour détecter les indicateurs de stress.

c) Méthodes basées sur l'analyse de l'activité :

Ces méthodes utilisent des données sur l'activité pour prédire le stress. Les données d'activité peuvent être collectées à partir de capteurs portables tels que les montres intelligentes, les capteurs de mouvement ou les capteurs de sommeil. Les données d'activité sont analysées en utilisant des techniques de traitement du signal et des algorithmes de machine Learning pour détecter les indicateurs de stress.

d) Méthodes basées sur les données comportementales :

Ces méthodes utilisent des données comportementales pour prédire le stress. Les données comportementales peuvent être collectées à partir de questionnaires, d'entretiens ou d'autres formes d'observation. Les données comportementales sont analysées en utilisant des techniques d'analyse de données et des algorithmes de machine Learning pour détecter les indicateurs de stress.

e) Méthodes basées sur l'analyse multimodale :

Ces méthodes combinent plusieurs sources de données pour prédire le stress. Par exemple, l'analyse multimodale peut combiner des données physiologiques et comportementales pour fournir une prédiction plus précise du stress.

II.3.2. Les techniques d'apprentissages automatiques utilisés dans la reconnaissance du stress :

Il existe plusieurs techniques d'apprentissage automatique utilisées pour la reconnaissance du stress, notamment :

a) **Les réseaux de neurones artificiels** :

Ces techniques utilisent des modèles de réseaux de neurones pour analyser les données d'entrée et prédire le niveau de stress.

b) **Les arbres de décision :**

Ces techniques utilisent des arbres de décision pour classifier les données en fonction du niveau de stress.

c) **Les machines à vecteurs de support (SVM) :**

Ces techniques utilisent des SVM pour classifier les données d'entrée en fonction du niveau de stress.

d) **Les réseaux de neurones convolutifs :**

Ces techniques utilisent des réseaux de neurones convolutifs pour extraire des caractéristiques à partir de données d'entrée telles que des images et prédire le niveau de stress.

e) **Les réseaux de neurones récurrents :**

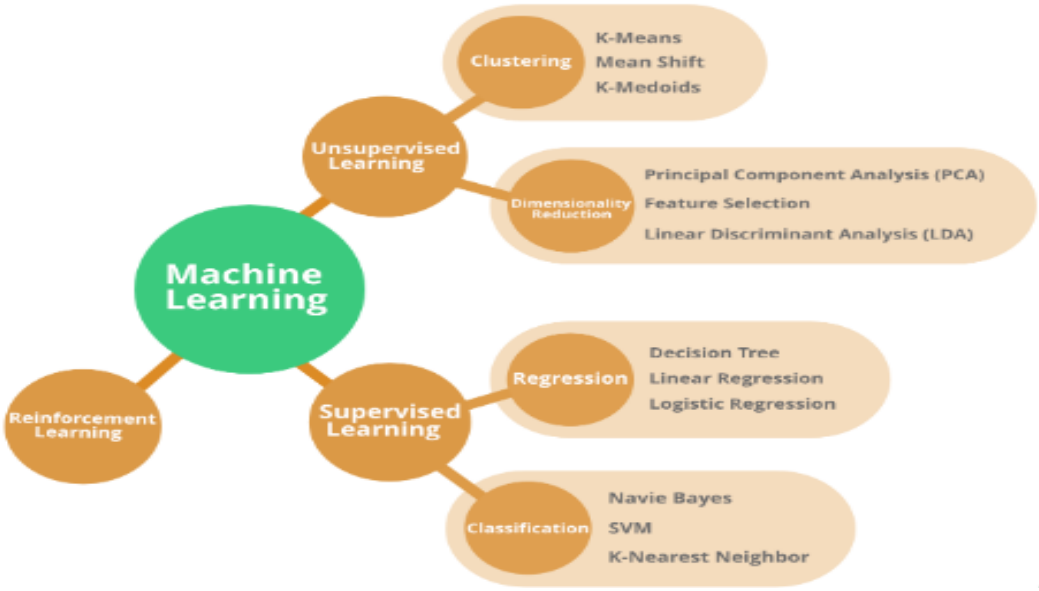
Ces techniques utilisent des réseaux de neurones récurrents pour modéliser des séquences de données, telles que des séquences de signaux physiologiques, et prédire le niveau de stress.

f) **Les méthodes de clustering :**

Ces techniques utilisent des algorithmes de clustering pour regrouper les données d'entrée en fonction de leur similarité, ce qui peut aider à identifier les modèles de stress.

g) les méthodes de régression linéaire

Les techniques mentionnées ci-dessus ne sont que quelques exemples parmi une grande variété de techniques disponibles pour la reconnaissance du stress.



II.4. Synthèse :

En résumé, la prédiction du stress est un domaine de recherche en pleine expansion qui utilise des méthodes d'apprentissage automatique pour détecter les indicateurs de stress dans les données physiologiques, linguistiques, d'activité et comportementales.

Il est important de noter que le choix de la technique d'apprentissage automatique dépendra de la nature des données disponibles et des objectifs de l'analyse.

Chapitre III : Choix des méthodes et Analyse

III.1. Introduction

Dans cette partie, nous présentons un aperçu des techniques d'apprentissages automatiques utilisés dans la reconnaissance du stress. Cette partie se divise en deux chapitres :

Le Chapitre 1 « Choix et Explication des méthodes» explicite les méthodes utilisés d'un point de vue de prédiction le stress, et montre les avantages et les inconvenants de chacun d'elles.

Le Chapitre 2 « Etude Comparative et simulations des méthodes » présente les différences et les simulations entre les différents méthodes choisies.

III.2. Choix et explication des méthodes :

Les méthodes utilisées dans le projet :

L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle qui fournit aux ordinateurs et aux systèmes informatiques la capacité d'apprendre et s'améliorer indépendamment de l'expérience précédente sans être explicitement programmé par un humain.

Ceci est très efficace dans le domaine de la santé car il y a d'énormes quantité de données, la prédiction résultante modèle sera inégalé et exempt d'erreurs humaines et réduit le temps nécessaire au diagnostic.

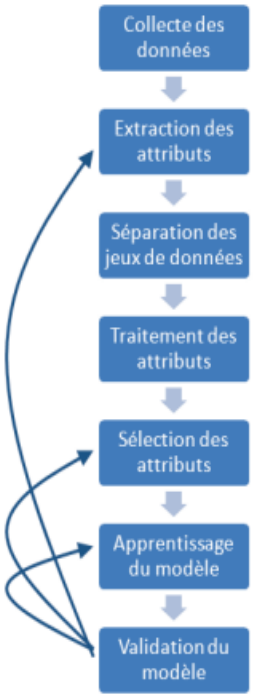
D'où nous avons utilisé 4 méthodes de notre état de l'art.

III.2.1. Régression linéaire (logistique)

Comme toutes les méthodes de régression, la régression logistique est une analyse prédictive. C'est utilisé dans les scénarios où une variable binaire et dépend d'une ou plusieurs variables indépendantes c.à.d. trouver une relation linéaire entre une variable cible et un ensemble de variables prédictives.

L'objectif de la régression linéaire est de trouver la ligne de régression qui minimise la distance entre les points de données réels et les prédictions faites par le modèle.

Ici, nous prenons les 7 attributs pertinents comme étant variables indépendantes et la possibilité d'un employé souffrant de stress et nécessitant un traitement comme variable dépendante qui doit être prédite par le modèle formé.



Voici les étapes de fonctionnement de l'algorithme de régression linéaire :

1. Collecte des données : Collecter les données sur la variable indépendante (X) et la variable dépendante (Y) pour l'ensemble d'entraînement.

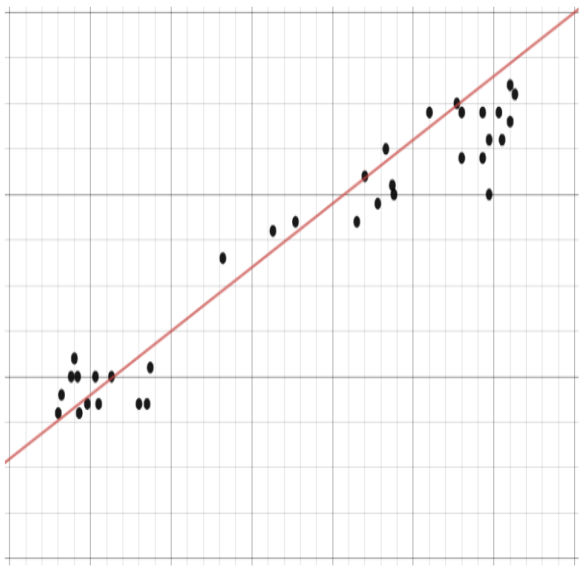
2. Visualisation des données : Tracer un graphique.

3. Séparation des données : Séparer l'ensemble de données en deux (l'ensemble d'entraînement et de test).

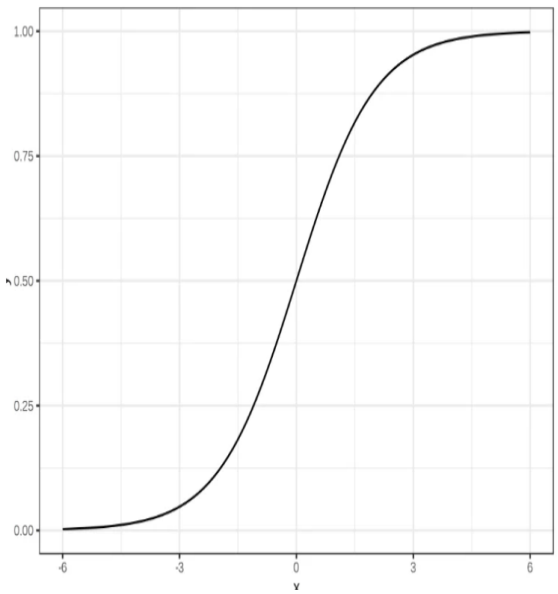
4. Entraînement du modèle : Utiliser l'ensemble d'entraînement pour ajuster les paramètres du modèle de régression linéaire. Les paramètres sont ajustés pour minimiser l'erreur quadratique moyenne (MSE) entre les valeurs prédites et les valeurs réelles.

5. Évaluation du modèle : Évaluer la performance du modèle en utilisant l'ensemble de test. On calcule l'erreur entre les valeurs prédites et les valeurs réelles de l'ensemble de test pour évaluer la qualité de la prédiction.

6. Utilisation du modèle : Utiliser le modèle pour prédire les valeurs de Y pour de nouvelles valeurs de X.



Exemple de régression linéaire



III.2.1.1. Avantages

Les avantages de l'algorithme de régression linéaire incluent :

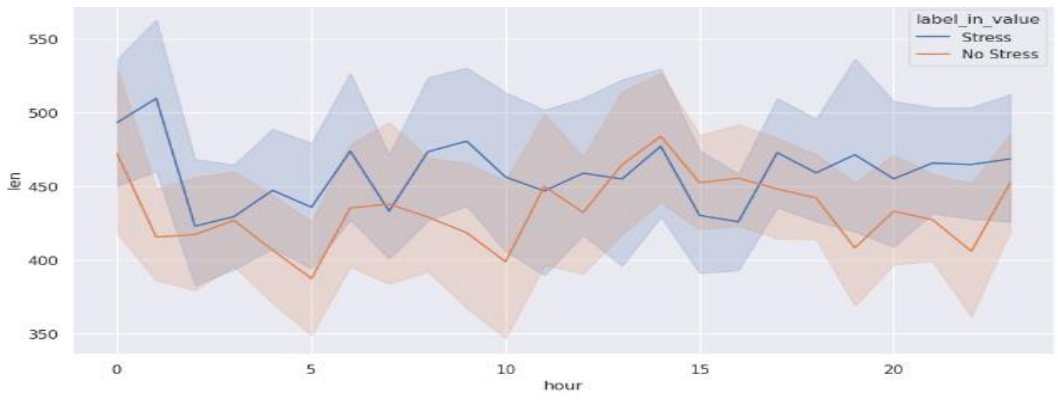
- **Simplicité :** C'est un algorithme simple et facile à comprendre. Il peut être implémenté facilement même par les débutants en apprentissage automatique.
- **Interopérabilité :** Les résultats de la régression linéaire peuvent être facilement interprétés, car l'effet de chaque variable sur la variable cible est quantifié par les coefficients de régression.
- **Rapidité :** C'est un algorithme rapide et efficace pour les ensembles de données de petite et moyenne taille.

III.2.1.2. Inconvénients

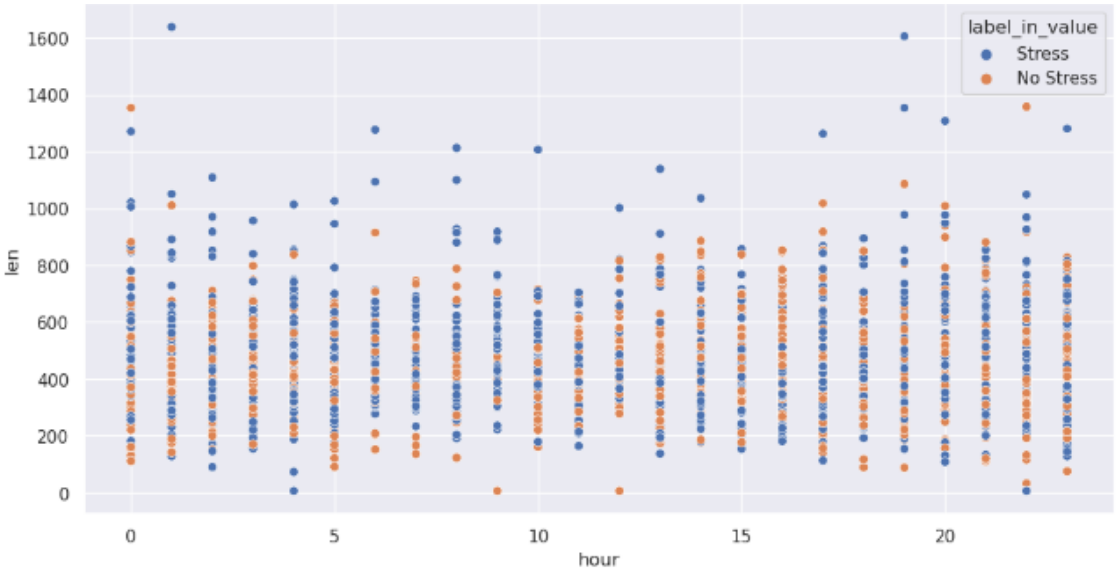
Cependant, l'algorithme de régression linéaire présente également des inconvénients :

- **Limitations de la linéarité :** La régression linéaire suppose une relation linéaire entre la variable cible et les variables prédictives. Si la relation est non linéaire, la performance de la régression linéaire peut être médiocre.
- **Sensibilité aux outliers :** Les outliers (valeurs aberrantes) peuvent avoir un impact important sur les résultats de la régression linéaire, car l'algorithme cherche à minimiser la distance entre les points de données réels et les prédictions.
- **Limitations de la sélection de variables :** Elle peut être sensible à la sélection des variables prédictives. Des variables non pertinentes ou redondantes peuvent affecter la qualité des prédictions.

En résumé, la régression linéaire est une méthode simple et efficace pour modéliser des relations linéaires entre des variables, mais elle peut présenter des limitations si les données ont des relations non linéaires, des valeurs aberrantes ou des variables non pertinentes.



Les textes longs sont un symptôme de stress.

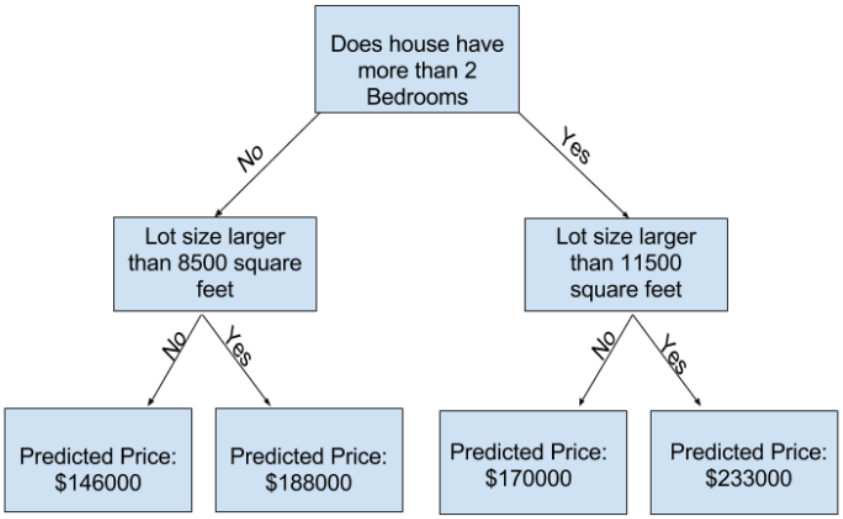


III.2.2. Arbre de décision

L'algorithme d'arbre de décision est une autre technique d'apprentissage automatique qui peut être utilisée pour prédire le stress humain. L'arbre de décision est un modèle d'apprentissage supervisé qui prend en entrée un ensemble d'observations et produit un arbre de décision hiérarchique, dans lequel chaque nœud représente une caractéristique de l'observation et chaque feuille représente une classe ou une décision, c.à.d. modéliser les choix multiples ou if-else déclarations/décisions sous forme d'arborescence.

Une fois que l'arbre de décision a été construit, il peut être utilisé pour prédire le niveau de stress pour de nouvelles observations. Cependant, comme pour la régression linéaire, il est important de disposer d'un ensemble de données bien annoté pour entraîner le modèle et il est important de valider le modèle sur des données de test pour évaluer ses performances.

Ici, des arbres de décision sont utilisés pour trouver le plus facteurs contributifs parmi les 7 caractéristiques qui sont utilisé. Ceci est très utile, car maintenant plus d'attention peuvent être donnés à ces zones et les mesures nécessaires sont prises sur ces lignes.



Exemple d'arbre de décision pris sur Kaggle

Voici quelques avantages et inconvénients de l'utilisation de l'algorithme d'arbre de décision :

III.2.2.1. Avantages

- Facile à comprendre et à interpréter
- Peut gérer des données manquantes ou des valeurs aberrantes
- Peut être utilisé pour des problèmes de classification et de régression
- Peut être utilisé avec des données qualitatives et quantitatives
- Peut être utilisé pour sélectionner les variables les plus importantes dans le modèle

III.2.2.2. Inconvénients

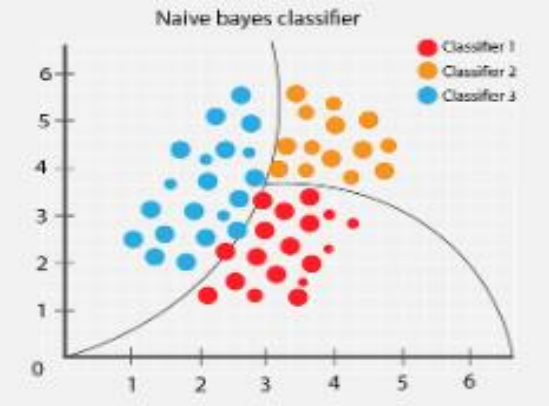
- Peut être sensible aux données d'entrée, ce qui peut entraîner un sur-ajustement ou un sous-ajustement du modèle
- Peut-être instable car de petits changements dans les données peuvent entraîner de grands changements dans l'arbre de décision
- Peut être limité par la complexité des modèles qu'il peut créer
- Peut nécessiter beaucoup de temps pour la construction de l'arbre de décision sur de grands ensembles de données

En somme, l'algorithme d'arbre de décision est une méthode puissante et largement utilisée pour la classification et la prédiction de variables, qui présente certains avantages et inconvénients qu'il est important de prendre en compte lors de son utilisation.

III.2.3. Naïve Bayes

L'algorithme **Naïve** Bayes est également utilisé pour la prédiction de stress dans les systèmes de machine Learning. C'est un algorithme d'apprentissage supervisé basé sur la probabilité et qui est particulièrement efficace pour la classification de textes.

Le fonctionnement de l'algorithme **Naïve Bayes** repose sur l'application du théorème de Bayes, qui consiste à calculer la probabilité d'un événement en connaissant les probabilités conditionnelles des événements qui le composent. Dans le cas de la prédiction de stress, l'algorithme de **Naïve Bayes** va chercher à prédire la probabilité que le stress soit présent ou non en fonction des différentes variables d'entrée.



$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

III.2.3.1. Avantages

- Peut être efficace pour la classification de textes
- Peut être utilisé pour de grandes bases de données
- Peut être utilisé avec des données manquantes
- Peut être utilisé avec des variables qualitatives et quantitatives

III.2.3.2. Inconvénients

- Peut être sensible aux variables corrélées, ce qui peut entraîner une faible précision de la prédiction
- Peut être limité par l'indépendance des variables, ce qui peut ne pas être le cas dans les données réelles
- Peut être difficile à interpréter en termes de contribution des variables à la prédiction
- Peut nécessiter des données d'entraînement importantes pour une prédiction précise

En somme, l'algorithme **Naïve** Bayes est une méthode puissante et largement utilisée pour la prédiction de stress, qui présente certains avantages et inconvénients qu'il est important de prendre en compte lors de son utilisation.

	precision	recall	f1-score	support
No Stress	0.76	0.57	0.65	335
Stress	0.69	0.84	0.76	375
accuracy			0.71	710
macro avg	0.72	0.70	0.70	710
weighted avg	0.72	0.71	0.71	710

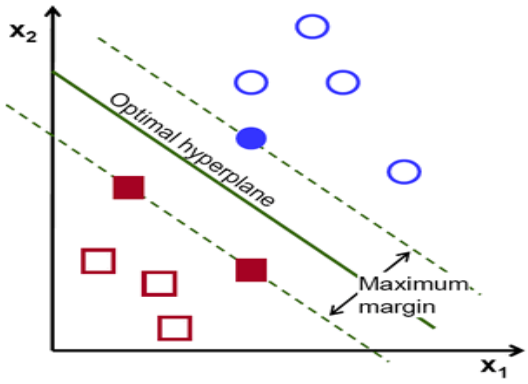
III.2.4. Les machines à vecteurs de support (SVM)

Les machines à vecteurs de support (SVM) sont un autre algorithme d'apprentissage automatique qui peut être utilisé pour la prédiction de stress. Voici quelques points sur leur fonctionnement :

- Les SVM sont une méthode de classification qui consiste à trouver l'hyperplan qui sépare le mieux les données en fonction des classes.

➔L'objectif des SVM est de trouver l'hyperplan qui maximise la marge entre les classes.

- Les SVM peuvent être très précis pour les données structurées et les données non linéaires à haute dimension et utilisés pour la classification binaire, mais ils peuvent également être étendus à la classification multi-classes.
- Les SVM peuvent également être utilisés pour la régression en utilisant une variante appelée machine à vecteurs de support de régression (SVR).



III.2.4.1. Avantages

Les avantages des machines à vecteurs de support (SVM) pour la prédiction de stress incluent :

- Les SVM ont une bonne précision de prédiction, en particulier pour les données structurées et les données non linéaires à haute dimension.
- Ils sont résistants aux valeurs aberrantes dans les données d'entraînement.
- Les SVM sont moins susceptibles de sur ajuster que certains autres algorithmes d'apprentissage automatique.

III.2.4.2. Inconvénients

Cependant, les SVM ont également des inconvénients potentiels, tels que :

- Les SVM peuvent être sensibles à la sélection des hyper paramètres, qui doivent être ajustés pour chaque ensemble de données spécifique.
- Etre coûteux en termes de temps de calcul et de mémoire lors de l'entraînement, en particulier pour les grands ensembles de données.
- Ne pas fonctionner correctement pour les données très bruyantes ou avec une grande proportion de chevauchements entre les classes.
- Les SVM ne fournissent pas directement des explications pour les prédictions effectuées, ce qui peut être un problème dans certains cas d'utilisation.

III.3. Etude comparative entre les méthodes

Voici une comparaison générale entre la régression linéaire, l'arbre de décision, Naïve Bayes et les machines à vecteurs de support pour la prédiction de stress :

Régression linéaire :

- Convient aux données linéaires et à une seule variable dépendante.
- Utilisé pour prédire des valeurs continues.
- Facile à comprendre et à interpréter.
- Convient pour les analyses de régression uni-variée.
- Peut être sensible aux valeurs aberrantes et à la présence de relations non linéaires.

Arbre de décision :

- Peut être utilisé pour les données catégorielles et continues.
- Convient pour les analyses de classification et de régression.
- Facile à comprendre et à visualiser.
- Peut être sensible aux valeurs aberrantes et au sur-apprentissage.

Naïve Bayes :

- Convient aux analyses de classification.
- Peut gérer les données avec de nombreuses caractéristiques.
- Peut être sensible aux hypothèses de base de la méthode Naïve Bayes.
- Estime les probabilités à partir des données d'entraînement.

Machines à vecteurs de support (SVM) :

- Convient aux données non linéaires et à haute dimension.
- Convient pour les analyses de classification et de régression.
- Peut gérer à la fois la classification binaire et la classification multi-classes.
- Peut être sensible à la sélection des hyper paramètres.
- Peut-être coûteux en termes de temps de calcul et de mémoire lors de l'entraînement.

En général, le choix de l'algorithme dépend de la nature des données, de l'objectif de la prédiction et des contraintes de temps et de ressources. Il est courant de comparer plusieurs algorithmes pour choisir celui qui convient le mieux à l'analyse des données.

==>Il n'y a pas de méthode unique qui convient à toutes les tâches de prédiction de stress humain, car cela dépendra des données spécifiques et des objectifs du modèle de prédiction. Chaque algorithme a ses avantages et ses inconvénients, et il est important de les considérer en fonction du contexte de l'application.

Par exemple, la régression linéaire peut être une bonne option si les données présentent une relation linéaire entre les variables et que la prédiction de stress doit être effectuée rapidement avec des ressources informatiques limitées. Les arbres de décision, quant à eux, peuvent être plus appropriés pour les données complexes avec des interactions entre les variables, car ils peuvent facilement capturer ces relations non linéaires.

Le choix de la méthode dépend également de la précision requise pour la prédiction. Dans certains cas, une méthode plus simple et moins précise peut suffire, tandis que dans d'autres cas, une méthode plus complexe et précise sera nécessaire. Il est donc important d'évaluer les performances de chaque algorithme sur les données spécifiques et de choisir celui qui convient le mieux aux besoins de l'application.

Conclusion générale

En conclusion, la prédiction du stress humain est un domaine de recherche en pleine expansion, qui suscite un intérêt croissant de la part de la communauté scientifique et des professionnels de la santé. Les algorithmes de machine Learning, tels que la régression linéaire, les arbres de décision, Naïve Bayes, les réseaux de neurones et les machines à vecteurs de support (SVM), sont souvent utilisés pour prédire le stress humain.

Chacun de ces algorithmes présente des avantages et des inconvénients en termes de précision de prédiction, de vitesse de traitement, de capacité d'interprétation des résultats, et de robustesse face à des données manquantes ou bruitées.

Il est donc important de bien comprendre les caractéristiques et les limitations de chaque algorithme, ainsi que les spécificités de chaque ensemble de données pour pouvoir choisir le meilleur algorithme pour la prédiction du stress humain.