Université Mohammed V

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes -Rabat- (ENSIAS)

End-of-year project report:

Prédiction de la COVID-19 à l'aide de Modèles de Machine Learning basés sur les Symptômes

---

Branch : Smart System Engineering (SSE)

Made by:
Yassin TALSSIS
Amine NASRI

Supervised by :
Pr. MESBAH Abderrahim

Academic year 2022-2023

Acknowledgement

Avant de commencer ce travail, nous voudrions d'abord exprimer nos remerciements à tous nos remerciements à tous ceux qui ont contribué, de près ou de loin, à l'élaboration et à la réalisation de ce travail.

Nous profitons de cette occasion pour exprimer notre gratitude à Monsieur ..... pour ses conseils, sa disponibilité et le partage constant de ses connaissances riches et variées.

Nous tenons également à remercier tous les enseignants qui nous ont aidés à trouver des solutions aux problèmes que nous avons rencontrés.

Nous tenons particulièrement à remercier ..... pour ses efforts continus pour tirer le maximum de nos capacités, ainsi que ..... pour nous avoir correctement orientés au sein du projet.

Enfin, nous tenons à remercier Monsieur ....., qui a assuré le suivi du projet, pour son soutien constant et ses encouragements.

Summary

Ce projet avait pour objectif de prédire la probabilité d'être atteint de la COVID-19 en utilisant des techniques de machine learning. Nous avons exploré différents modèles tels que la régression linéaire, KNN, Naive Bayes et les réseaux neuronaux pour identifier les indicateurs les plus significatifs et prendre des décisions éclairées en matière de santé publique et de prévention. Les résultats obtenus ont montré que ces modèles peuvent fournir des informations précieuses pour surveiller les symptômes, gérer les ressources médicales et identifier les populations à risque élevé. Ce projet a souligné l'importance de l'apprentissage automatique dans la lutte contre la COVID-19 et ouvre la voie à de nouvelles recherches et applications dans ce domaine critique de la santé publique.

Abstract

The objective of this project was to predict the likelihood of being affected by COVID-19 using machine learning techniques. Various models including linear regression, KNN, Naive Bayes, and neural networks were explored to identify the most significant indicators and make informed decisions in public health and prevention. The results demonstrated that these models can provide valuable insights for monitoring symptoms, managing medical resources, and identifying high-risk populations. This project highlights the importance of machine learning in combating COVID-19 and paves the way for further research and applications in this critical field of public health.

Table of contents

List of Figures

List of Tables

General Introduction

La pandémie de COVID-19 a eu un impact majeur sur le monde entier, mettant en évidence l'importance de la prévention, de la détection précoce et de la gestion efficace de cette maladie. Dans ce contexte, une question cruciale se pose : quels sont les indicateurs les plus significatifs pour prédire si une personne est susceptible d'être atteinte de la COVID-19 ? La réponse à cette question revêt une grande importance, car elle peut nous aider à identifier les principaux symptômes à surveiller, les facteurs de risque les plus importants et à prendre des décisions éclairées en matière de santé publique et de prévention.

L'identification des indicateurs les plus significatifs pour prédire la présence de la COVID-19 peut fournir des informations essentielles pour le développement de mesures préventives ciblées. En comprenant les symptômes les plus prédictifs, les professionnels de la santé peuvent mettre en place des stratégies de dépistage plus efficaces, permettant une détection précoce des cas suspects et une isolation rapide des individus atteints.

De plus, en identifiant les facteurs de risque les plus importants, nous pouvons mieux comprendre les populations à risque élevé et prendre des mesures spécifiques pour les protéger. Cela peut inclure la mise en place de programmes de vaccination prioritaires, l'adaptation des protocoles de traitement et la gestion appropriée des ressources médicales pour répondre aux besoins des personnes les plus vulnérables.

Dans cette étude, nous explorons différentes approches, notamment l'utilisation de techniques de machine learning, pour déterminer quels sont les indicateurs les plus significatifs pour prédire la susceptibilité à la COVID-19. Nous examinons une gamme de symptômes couramment associés à la maladie, ainsi que d'autres variables pertinentes, telles que l'âge, le sexe, les antécédents médicaux, etc. En analysant ces données, nous visons à fournir des informations précieuses pour la prise de décisions éclairées en matière de santé publique, de prévention et de gestion des ressources médicales.

En résumé, cette étude vise à répondre à la problématique essentielle de déterminer les indicateurs les plus significatifs pour prédire la susceptibilité à la COVID-19. En identifiant ces indicateurs clés, nous pouvons renforcer nos capacités de prévention, de détection précoce et de gestion de cette maladie, contribuant ainsi à la santé publique mondiale et à la protection des populations à risque élevé.

Chapter 1

Problematique and objectifs

1.1 Introduction

Dans ce chapitre, nous aborderons la problématique de déterminer les indicateurs les plus significatifs pour prédire si une personne est susceptible d'être atteinte de la COVID-19. Nous explorerons les principaux objectifs de notre étude, qui sont de mieux comprendre les symptômes et les facteurs de risque liés à cette maladie, ainsi que de fournir des informations précieuses pour la prise de décisions en matière de santé publique et de prévention.

1.2 Problematique

Notre problimatique est de chercher à déterminer quels sont les indicateurs les plus significatifs pour prédire si une personne est susceptible d'être atteinte de la COVID-19. Cela pourrait nous aider à identifier les principaux symptômes à surveiller, les facteurs de risque les plus importants et à prendre des décisions éclairées en matière de santé publique et de prévention, tels que la mise en place de mesures préventives ciblées, la gestion des ressources médicales ou l'identification des populations à risque élevé.

1.3 Objectifs :

Dans ce contexte, les principaux objectifs de notre étude sont les suivants :

- Identifier les indicateurs les plus significatifs : Nous chercherons à déterminer quels symptômes, caractéristiques et facteurs de risque sont les plus prédictifs de la COVID-19. En analysant les données disponibles, nous chercherons à trouver des modèles et des relations significatives qui peuvent aider à prédire la susceptibilité à cette maladie.
- Comprendre les principaux symptômes à surveiller : En identifiant les symptômes les plus couramment associés à la COVID-19, nous pourrons mieux informer les professionnels de la santé et le grand public sur les signes précurseurs à surveiller. Cela permettra une détection précoce des cas suspects et une prise en charge appropriée pour réduire la transmission du virus.
- Contribuer à la santé publique et à la prévention : En fournissant des informations sur les indicateurs significatifs, notre étude vise à soutenir la prise de décisions en matière de santé publique. Cela inclut la mise en place de mesures préventives ciblées, la gestion des ressources médicales et l'identification des populations à risque élevé afin de mieux protéger les individus et les communautés.

1.4 Solution proposée

Afin de résoudre cette problématique, nous allons utiliser des techniques de machine learning pour déterminer si une personne est atteinte de la COVID-19 en se basant sur des indicateurs tels que les problèmes respiratoires, la fièvre, les maux de tête, etc. Le modèle de machine learning entraîné sera utilisé pour prédire la probabilité qu'une personne soit atteinte de la COVID-19 en fonction de ces symptômes. Cette approche peut fournir des informations précieuses pour prendre des décisions éclairées en matière de santé publique, de prévention et de gestion des ressources médicales.

Chapter 2

Étude technique

2.1 Les outils utilisés

La mise en œuvre de notre solution a impliqué plusieurs outils clés, chacun apportant des capacités uniques à notre processus de développement.

- **TensorFlow:** TensorFlow est une plate-forme open source pour l'apprentissage automatique. Il fournit des bibliothèques et des ressources complètes qui nous ont permis de former notre modèle d'apprentissage en profondeur pour la classification de la maturité des tomates.

TensorFlow Logo

- **Scikit-learn:** Scikit-learn (sklearn) est une bibliothèque populaire d'apprentissage automatique en Python, offrant une large gamme d'outils et d'algorithmes pour la préparation des données et la construction de modèles prédictifs. Elle est largement utilisée pour sa simplicité d'utilisation et sa polyvalence dans différents domaines de l'apprentissage automatique.

Scikit-learn Logo

- **Kaggle:** Kaggle is an online community of data scientists and machine learners, owned by Google. We used Kaggle to access a good development environment.

Kaggle Logo

Ces outils, chacun ayant un objectif distinct, ont été combinés pour former la base technique de notre solution visant à prédire la COVID-19.

2.2 Description de base de données

Cette base de données contient des informations relatives à plusieurs symptômes et conditions médicales, ainsi que des facteurs de risque liés à la COVID-19. Les attributs incluent des indications sur des problèmes respiratoires, de la fièvre, de la toux sèche, des maux de gorge, un écoulement nasal, de l'asthme, des maladies pulmonaires chroniques, des maux de tête, des maladies cardiaques, du diabète, de la fatigue, des problèmes gastro-intestinaux, des voyages à l'étranger, des contacts avec des patients atteints de la COVID-19, la participation à de grands rassemblements, des visites de lieux publics exposés, des membres de la famille travaillant dans des endroits exposés au public, le port de masques, la désinfection provenant du marché, ainsi que la présence de la COVID-19. Cette base de données permet

d'analyser et de suivre les facteurs de risque associés à la COVID-19 et de prendre des décisions éclairées en matière de santé publique et de prévention.

	Breathing Problem	Fever	Dry Cough	Sore throat	Running Nose	Asthma	Chronic Lung Disease	Headache	Heart Disease	Diabetes	...	Fatigue	Gastrointestinal	Abroad travel	Contact with COVID Patient	Attended Large Gathering	Visited Public Exposed Places	Family working in Public Exposed Places	Wearing Masks	Sanitization from Market	COVID-19
0	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	...	Yes	Yes	No	Yes	No	Yes	Yes	No	No	Yes
1	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	...	Yes	No	No	No	Yes	Yes	No	No	No	Yes
2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	...	Yes	Yes	Yes	No	No	No	No	No	No	Yes
3	Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes	...	No	No	Yes	No	Yes	Yes	No	No	No	Yes
4	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	...	No	Yes	No	Yes	No	Yes	No	No	No	Yes

Les colonnes de la base de données

2.3 Préparation des données

Étant donné que les valeurs que nous avons sont catégorielles, nous allons utiliser la classe LabelEncoder pour convertir ces labels non numériques en labels numériques. Le LabelEncoder va assigner un nombre unique à chaque catégorie, facilitant ainsi l'analyse et l'utilisation ultérieure des données dans les algorithmes d'apprentissage automatique.

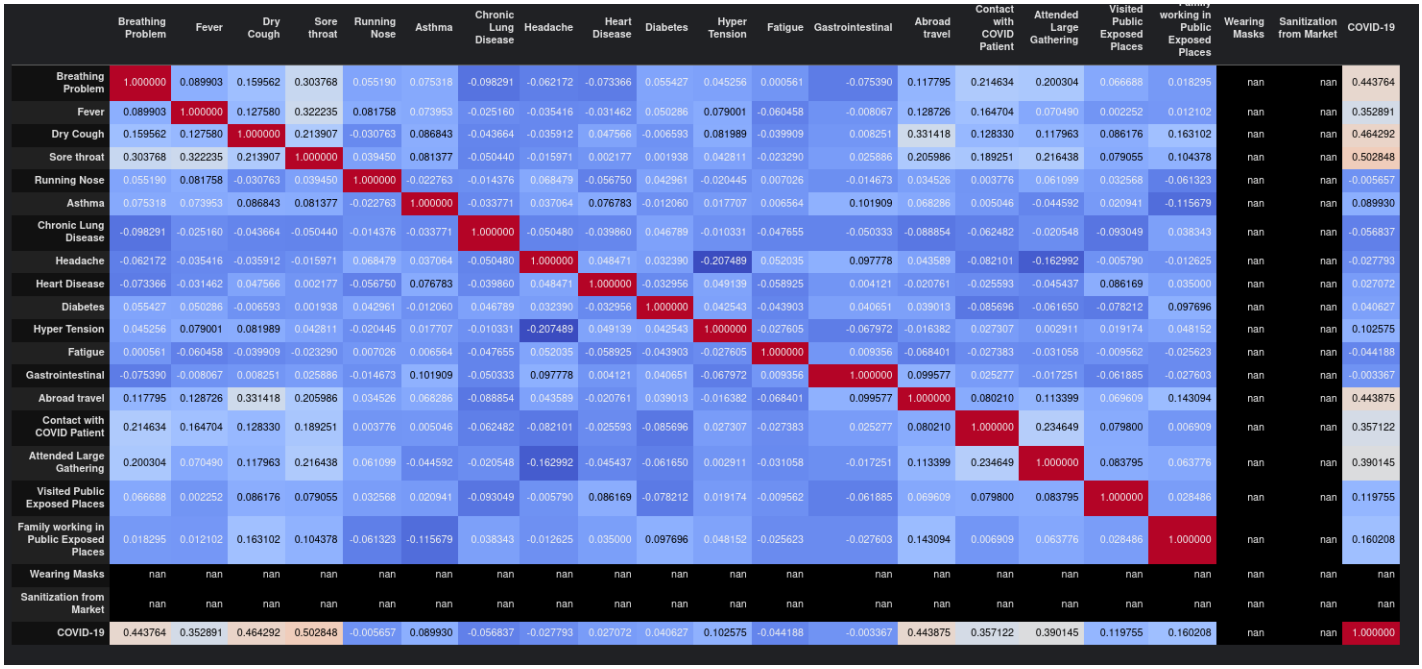
	Breathing Problem	Fever	Dry Cough	Sore throat	Running Nose	Asthma	Chronic Lung Disease	Headache	Heart Disease	Diabetes	...	Fatigue	Gastrointestinal	Abroad travel	Contact with COVID Patient	Attended Large Gathering	Visited Public Exposed Places	Family working in Public Exposed Places	Wearing Masks	Sanitization from Market	COVID-19
0	1	1	1	1	1	0	0	0	0	1	...	1	1	0	1	0	1	1	0	0	1
1	1	1	1	1	0	1	1	1	0	0	...	1	0	0	0	1	1	0	0	0	1
2	1	1	1	1	1	1	1	1	0	1	...	1	1	1	0	0	0	0	0	0	1
3	1	1	1	0	0	1	0	0	1	1	...	0	0	1	0	1	1	0	0	0	1
4	1	1	1	1	1	0	1	1	1	1	...	0	1	0	1	0	1	0	0	0	1

Les colonnes de la base de données

2.4 Vérifier la dépendance des variables

Après avoir analysé les données, nous avons identifié certaines variables qui présentent une faible corrélation avec la variable cible ('COVID-19'). Par conséquent, nous allons les éliminer de notre ensemble de données afin de construire un modèle de

prédiction plus précis. Cette décision est basée sur notre constatation d'une forte corrélation entre les variables restantes et la variable cible, ce qui indique qu'elles sont plus étroitement liées à la prédiction de la présence de la COVID-19. En supprimant les variables non pertinentes, nous améliorons la qualité de nos données et augmentons ainsi les chances d'obtenir des prédictions plus fiables et précises.Voir la figure suivante



	Breathing Problem	Fever	Dry Cough	Sore throat	Running Nose	Asthma	Chronic Lung Disease	Headache	Heart Disease	Diabetes	Hyper Tension	Fatigue	Gastrointestinal	Abroad travel	Contact with COVID Patient	Attended Large Gathering	Visited Public Exposed Places	Family working in Public Exposed Places	Wearing Masks	Sanitization from Market	COVID-19
Breathing Problem	1.000000	0.089903	0.159562	0.303768	0.055190	0.075318	-0.098291	-0.062172	-0.073366	0.055427	0.045256	0.000561	-0.075390	0.117795	0.214634	0.200304	0.066688	0.018295	nan	nan	0.443764
Fever	0.089903	1.000000	0.127580	0.322235	0.081758	0.073953	-0.025160	-0.035416	-0.031462	0.050286	0.079001	-0.060458	-0.008067	0.128726	0.164704	0.070490	0.002252	0.012102	nan	nan	0.352891
Dry Cough	0.159562	0.127580	1.000000	0.213907	-0.030763	0.086843	-0.043664	-0.035912	0.047566	-0.006593	0.081989	-0.039909	0.008251	0.331418	0.128330	0.117963	0.086176	0.163102	nan	nan	0.464292
Sore throat	0.303768	0.322235	0.213907	1.000000	0.039450	0.081377	-0.050440	-0.015971	0.002177	0.001938	0.042811	-0.023290	0.025886	0.205986	0.189251	0.216438	0.079055	0.104378	nan	nan	0.502848
Running Nose	0.055190	0.081758	-0.030763	0.039450	1.000000	-0.022763	-0.014376	0.068479	-0.056750	0.042961	-0.020445	0.007026	-0.014673	0.034526	0.003776	0.061099	0.032568	-0.061323	nan	nan	-0.005657
Asthma	0.075318	0.073953	0.086843	0.081377	-0.022763	1.000000	-0.033771	0.037064	0.076783	-0.012060	0.017707	0.006564	0.101909	0.068286	0.005046	-0.044592	0.020941	-0.115679	nan	nan	0.089930
Chronic Lung Disease	-0.098291	-0.025160	-0.043664	-0.050440	-0.014376	-0.033771	1.000000	-0.050480	-0.039860	0.046789	-0.010331	-0.047655	-0.050333	-0.088854	-0.062482	-0.020548	-0.093049	0.038343	nan	nan	-0.056837
Headache	-0.062172	-0.035416	-0.035912	-0.015971	0.068479	0.037064	-0.050480	1.000000	0.048471	0.032390	-0.207489	0.052035	0.097778	0.043589	-0.082101	-0.162992	-0.005790	-0.012625	nan	nan	-0.027793
Heart Disease	-0.073366	-0.031462	0.047566	0.002177	-0.056750	0.076783	-0.039860	0.048471	1.000000	-0.032956	0.049139	-0.058925	0.004121	-0.020761	-0.025593	-0.045437	0.086169	0.035000	nan	nan	0.027072
Diabetes	0.055427	0.050286	-0.006593	0.001938	0.042961	-0.012060	0.046789	0.032390	-0.032956	1.000000	0.042543	-0.043903	0.040651	0.039013	-0.085696	-0.061650	-0.078212	0.097696	nan	nan	0.040627
Hyper Tension	0.045256	0.079001	0.081989	0.042811	-0.020445	0.017707	-0.010331	-0.207489	0.049139	0.042543	1.000000	-0.027605	-0.067972	-0.016382	0.027307	0.002911	0.019174	0.048152	nan	nan	0.102575
Fatigue	0.000561	-0.060458	-0.039909	-0.023290	0.007026	0.006564	-0.047655	0.052035	-0.058925	-0.043903	-0.027605	1.000000	0.009356	-0.068401	-0.027383	-0.031058	-0.009562	-0.025623	nan	nan	-0.044188
Gastrointestinal	-0.075390	-0.008067	0.008251	0.025886	-0.014673	0.101909	-0.050333	0.097778	0.004121	0.040651	-0.067972	0.009356	1.000000	0.099577	0.025277	-0.017251	-0.061885	-0.027603	nan	nan	-0.003367
Abroad travel	0.117795	0.128726	0.331418	0.205986	0.034526	0.068286	-0.088854	0.043589	-0.020761	0.039013	-0.016382	-0.068401	0.099577	1.000000	0.080210	0.113399	0.069609	0.143094	nan	nan	0.443875
Contact with COVID Patient	0.214634	0.164704	0.128330	0.189251	0.003776	0.005046	-0.062482	-0.082101	-0.025593	-0.085696	0.027307	-0.027383	0.025277	0.080210	1.000000	0.234649	0.079800	0.006909	nan	nan	0.357122
Attended Large Gathering	0.200304	0.070490	0.117963	0.216438	0.061099	-0.044592	-0.020548	-0.162992	-0.045437	-0.061650	0.002911	-0.031058	-0.017251	0.113399	0.234649	1.000000	0.083795	0.063776	nan	nan	0.390145
Visited Public Exposed Places	0.066688	0.002252	0.086176	0.079055	0.032568	0.020941	-0.093049	-0.005790	0.086169	-0.078212	0.019174	-0.009562	-0.061885	0.069609	0.079800	0.083795	1.000000	0.028486	nan	nan	0.119755
Family working in Public Exposed Places	0.018295	0.012102	0.163102	0.104378	-0.061323	-0.115679	0.038343	-0.012625	0.035000	0.097696	0.048152	-0.025623	-0.027603	0.143094	0.006909	0.063776	0.028486	1.000000	nan	nan	0.160208
Wearing Masks	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
Sanitization from Market	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
COVID-19	0.443764	0.352891	0.464292	0.502848	-0.005657	0.089930	-0.056837	-0.027793	0.027072	0.040627	0.102575	-0.044188	-0.003367	0.443875	0.357122	0.390145	0.119755	0.160208	nan	nan	1.000000

Matrice de corrélation

2.5 Prétraitement des données

Cette étape a pour objectif de supprimer les colonnes qui ne contiennent pas d'informations utiles ou qui ont une corrélation très faible avec la variable cible (dans ce cas, la COVID-19). Cette approche permet d'améliorer la qualité des données et d'augmenter les chances d'obtenir de meilleurs résultats lors de la prédiction de la COVID-19.

	Breathing Problem	Fever	Dry Cough	Sore throat	Hyper Tension	Abroad travel	Contact with COVID Patient	Attended Large Gathering	Visited Public Exposed Places	Family working in Public Exposed Places	COVID-19
Breathing Problem	1.000000	0.089903	0.159562	0.303768	0.045256	0.117795	0.214634	0.200304	0.066688	0.018295	0.443764
Fever	0.089903	1.000000	0.127580	0.322235	0.079001	0.128726	0.164704	0.070490	0.002252	0.012102	0.352891
Dry Cough	0.159562	0.127580	1.000000	0.213907	0.081989	0.331418	0.128330	0.117963	0.086176	0.163102	0.464292
Sore throat	0.303768	0.322235	0.213907	1.000000	0.042811	0.205986	0.189251	0.216438	0.079055	0.104378	0.502848
Hyper Tension	0.045256	0.079001	0.081989	0.042811	1.000000	-0.016382	0.027307	0.002911	0.019174	0.048152	0.102575
Abroad travel	0.117795	0.128726	0.331418	0.205986	-0.016382	1.000000	0.080210	0.113399	0.069609	0.143094	0.443875
Contact with COVID Patient	0.214634	0.164704	0.128330	0.189251	0.027307	0.080210	1.000000	0.234649	0.079800	0.006909	0.357122
Attended Large Gathering	0.200304	0.070490	0.117963	0.216438	0.002911	0.113399	0.234649	1.000000	0.083795	0.063776	0.390145
Visited Public Exposed Places	0.066688	0.002252	0.086176	0.079055	0.019174	0.069609	0.079800	0.083795	1.000000	0.028486	0.119755
Family working in Public Exposed Places	0.018295	0.012102	0.163102	0.104378	0.048152	0.143094	0.006909	0.063776	0.028486	1.000000	0.160208
COVID-19	0.443764	0.352891	0.464292	0.502848	0.102575	0.443875	0.357122	0.390145	0.119755	0.160208	1.000000

Matrice de corrélation

2.6 Les algorithmes utilisé

Dans le cadre de la résolution de ce problème, plusieurs modèles de machine learning sont utilisés pour prédire si une personne est susceptible d'être atteinte de la COVID-19. Voici quelques exemples de ces modèles :

- Régression linéaire: La régression linéaire est un modèle simple qui tente de trouver une relation linéaire entre les variables d'entrée et la variable cible. Il est couramment utilisé pour estimer des valeurs continues et peut être adapté pour prédire la probabilité d'être atteint de la COVID-19 en fonction des symptômes et des facteurs de risque.
- K plus proches voisins (KNN) : Le modèle KNN est basé sur le principe que des observations similaires se regroupent dans l'espace. Il calcule la similarité entre les nouvelles données et les données d'entraînement, puis attribue une étiquette en fonction des classes majoritaires des voisins les plus proches. Dans notre cas, les indicateurs tels que les problèmes de respiration, la fièvre, les maux de tête, etc., sont utilisés pour évaluer la similitude et prédire la présence de la COVID-19.
- Naive Bayes : Le modèle Naive Bayes est basé sur le théorème de Bayes et suppose que les caractéristiques sont indépendantes les unes des autres. Il

estime la probabilité d'appartenance à chaque classe en utilisant les probabilités conditionnelles des caractéristiques. En utilisant les indicateurs liés à la COVID-19, ce modèle peut calculer la probabilité qu'une personne soit atteinte de la maladie.

- Réseaux de neurones (Neural Networks) : Les réseaux de neurones sont des modèles d'apprentissage profond capables de capturer des relations complexes entre les caractéristiques et les étiquettes. Ils sont constitués de plusieurs couches de neurones interconnectés. En utilisant des architectures appropriées, les réseaux de neurones peuvent être entraînés sur les indicateurs tels que les problèmes de respiration, la fièvre, les maux de tête, etc., pour prédire la présence de la COVID-19.

Ces différents modèles de machine learning offrent des approches variées pour résoudre le problème de prédiction de la COVID-19. Chacun a ses avantages et ses limitations, et leur performance peut varier en fonction de la qualité des données, de la taille de l'ensemble d'entraînement et des paramètres choisis. Il est important d'expérimenter avec plusieurs modèles et d'évaluer leurs performances pour trouver celui qui convient le mieux à la résolution de cette problématique spécifique.

2.7 Results and evaluation

En fonction des valeurs d'exactitude (accuracy) fournies pour chaque modèle, voici une comparaison des performances :

L'ANN (Artificial Neural Network) a obtenu la plus haute précision avec une accuracy de 0.97 et une loss (perte) de 0,04. Cela indique que les réseaux de neurones ont

Architecture	Regression Lineare	KNN	Naive Bayes	ANN
Training Accuracy	0.96	0.97	0.74	0.97
Testing Accuracy	0.96	0.96	0.76	0.76

Table 2.1: Comparaison des résultats

réussi à capturer des relations complexes entre les caractéristiques et les étiquettes, conduisant à une prédiction très précise.

Le KNN (K-Nearest Neighbors) et la régression logistique ont également montré de bonnes performances avec des accuracies respectives de 0,96 et 0,95. Ces modèles ont démontré leur capacité à prédire avec précision la présence de la COVID-19 en utilisant les indicateurs sélectionnés. Cependant, le modèle Naive Bayes a affiché une accuracy plus basse de 0.76, ce qui souligne les limitations de l'hypothèse d'indépendance des caractéristiques dans ce contexte spécifique.

2.8 Conclusion

En fin de compte, le choix du modèle de machine learning dépendra des objectifs spécifiques, des caractéristiques des données et des ressources disponibles. Des expérimentations supplémentaires et une validation rigoureuse sont nécessaires pour confirmer et affiner les performances des modèles et leur applicabilité dans le contexte de prédiction de la COVID-19.

Conclusion générale

En conclusion, ce projet visant à prédire la probabilité d'être atteint de la COVID-19 en se basant sur des indicateurs spécifiques a été réalisé en utilisant des techniques de machine learning telles que la régression linéaire, KNN, Naive Bayes et les réseaux neuronaux. Les résultats obtenus démontrent que ces modèles sont capables d'apporter des informations précieuses pour identifier les principaux symptômes à surveiller et les facteurs de risque les plus importants.

Grâce à l'analyse des données et à l'utilisation de ces modèles, nous pouvons prendre des décisions éclairées en matière de santé publique et de prévention. Ces décisions peuvent inclure la mise en place de mesures préventives ciblées, la gestion efficace des ressources médicales et l'identification des populations à risque élevé.

En somme, ce projet met en évidence l'importance de l'utilisation des techniques de machine learning dans la prédiction de la COVID-19. Ces modèles peuvent aider à identifier les indicateurs les plus significatifs, à renforcer les mesures de santé publique et à prendre des décisions éclairées pour la prévention et la gestion de cette pandémie mondiale.

Bibliography

https://www.edureka.co/blog/covid-19-outbreak-prediction-using-machine-learning/

https://www.edureka.co/blog/covid-19-outbreak-prediction-using-machine-learning/

https://st3824.medium.com/predicting-covid-19-cases-using-machine-learning-5bf0d6990eef

https://medium.com/applications-of-machine-learning/predicting-icu-admission-of-confirmed-covid-19-patients-using-machine-learning-algorithm-4a4f41e21d8b