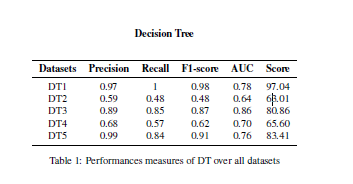
DT.

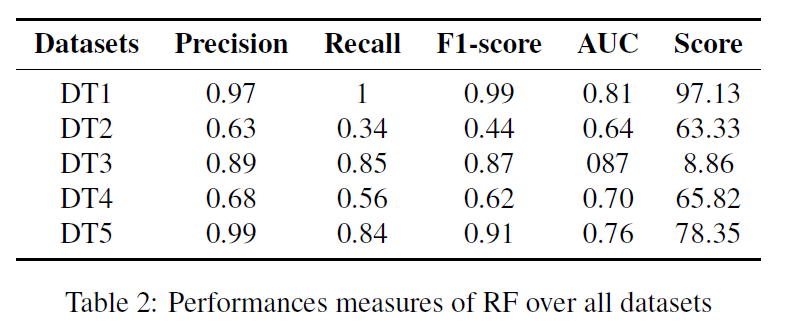
Le tableau 1 montre les mesures de performance (précision, rappel, mesure F et précision) des résultats de notre classificateur DT après expérimentation sur tous nos jeux de données. L’observation montre que le classifier DT présente les meilleurs score sur les jeux de données DT1, DT3 et DT5 avec des scores de 97.04%, 80.86% et 83.41% respectivement. On retrouve également les plus grandes valeurs des AUC (Area Under the Curve) pour ces mêmes jeux de données avec 0.78, 0.86 et 0.76 respectivement. Cependant on note que les valeurs de la sensibilité sont plus élevées que les valeurs de la spécificité sur les jeux de données DT1 et DT5 et qu’ils sont sensiblement identiques pour les jeux de données DT2, DT3 et DT4. Ceci signifie que la DT est plus enclin à prédire aussi bien si patient donné est atteint de paludisme ou s’il ne l’est pas sur les jeux de donnés DT2, DT3 et DT4 contrairement sur les jeux de donnés DT1 et DT5 ou notre classifieur est performant uniquement pour prédire si un patient donné est atteint de paludisme. Cette même tendance est observée sur les F-scores des valeurs variant de 0.91 à 0.98 sur les jeux de données DT1, DT3 et DT5.

Table 1 shows the performance measures (precision, recall, F measure and precision) of the results of our Decision Tree classifier after experimentation on all our datasets. The observation shows that the best scores of our classifier are achieved on the datasets DT1, DT3 and DT5 which are 7.04%, 80.86% and 83.41% respectively. Also AUC (Area Under the Curve) values are higher ​​for these same datasets which are 0.78, 0.86 and 0.76 respectively. However, we note that the sensitivity values ​​are higher than the specificity values ​​on the datasets DT1 and DT5 so that they are substantially identical for the datasets DT2, DT3 and DT4. This means that DT is more inclined to predict as well whether a given patient has malaria or he doesn’t, on the datasets DT2, DT3 and DT4, while our classifier on the datasets DT1 and DT5 our classifier is only efficient in predicting whether a given patient has malaria. This same trend is observed on the F-scores which higher values varying between 0.91 and 0.98 on the datasets DT1, DT3 and DT5.



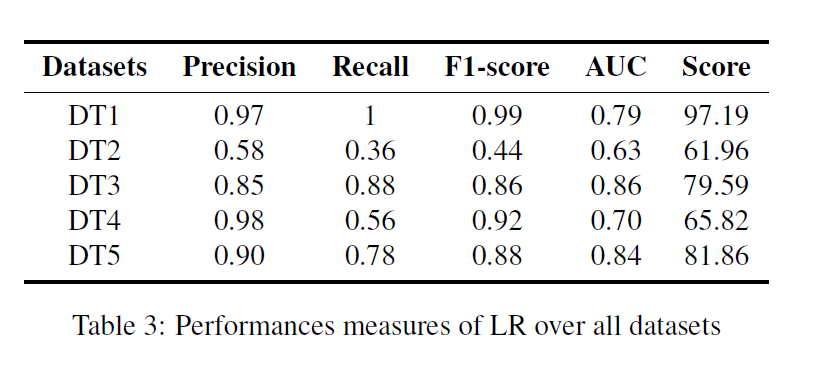
**Random Forest**

The performance of the random forest varied throughout the study depending on the dataset, although overall it performed well as shown in Table 2. Notice that best accuracy are achieved by random forest classifier on the datasets DT1, DT2 and DT5 which are respectively 97.13%, 80.86% and 78.35. In contrast with the results obtained with the DT classifier, the Sensivity values are higher than specificity values on datasets DT1 and DT5 whereas the inverse is noticed on the dataset DT3. At the same time we note that these values are roughly identical on the datasets DT3 and D4.



**Logistic regression**

In table 3 we show the performance measures LR classifier experimented on our five datasets. We notice that our classifier have overall precision which are vary between 58% and 98%.



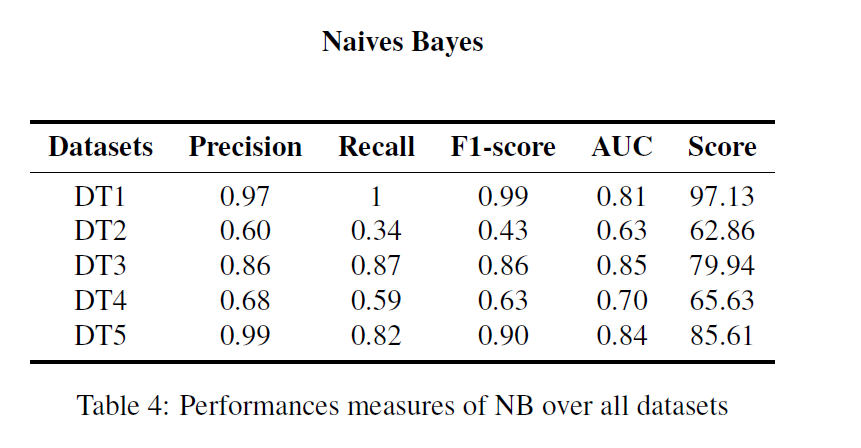
We observe that the higher precision is obtained with DT4 dataset while the corresponding score is equal to 65.82% is the lowest of all other datasets. Also we notice that the LR presents homogeneous results on the DT3 dataset with an accuracy of 85%, a sensitivity equal to 88%, an F-score of 92%, an AUC which is 0.86 and a score equal to 79.59%. . We also note that the best AUC and the best F-score are obtained by LR on the DT3 dataset.

Nous observons que la plus grande précision est obtenue avec le jeu de données DT4 alors que le score correspondant est presque le plus bas des tous les autres. Egalement on remarque que la LR présente des résultats homogènes sur le jeu de données DT3 avec une précision de 85%, une sensibilité égale à 88%, un F-score de 92%, un AUC qui est 0.86 et un score de 79.59%. On note Aussi que le meilleur AUC et le meilleur F-score sont obtenues par LR sur le jeu de données DT3

**Naives Bayes**

Contrairement aux résultats ci-dessus le classificateur NB présente des performances très hétérogènes en fonction des mesures de performances utilisées. En effet on observe la meilleure précision sur le jeu de données DT5 (99%), le meilleur F-score est quant à lui obtenu sur le jeu de données DT1 (0.99), la meilleure AUC sur DT3 (0.85) et le meilleur score sur le jeu de données DT1. On notera aussi que la meilleure spécificité est obtenue sur DT4 et varie entre 0.65 et 0.70 (voir annexes).

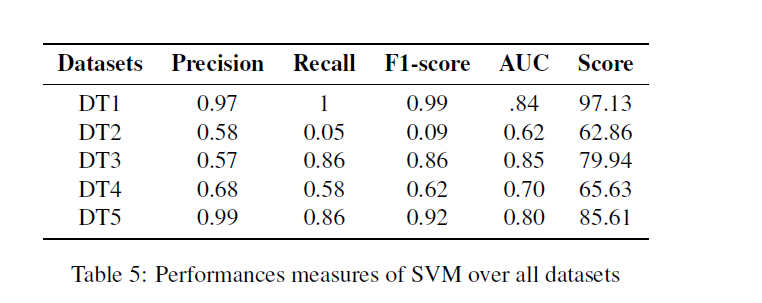
In contrast with the results above, NB classifier presents very heterogeneous performances regarding the performance measures used. In fact, we observe that the best precision is achieved on the dataset DT5 which is 99%, although the best F-score and the higher accuracy are obtained on the dataset DT1 which are 0.99 and 97.13% respectively and finally the best AUC is observed on the dataset DT3 which is 0.85. We also note that the best specificity is obtained on DT4 and varies between 0.65 and 0.70 (see appendices).



**SVM**

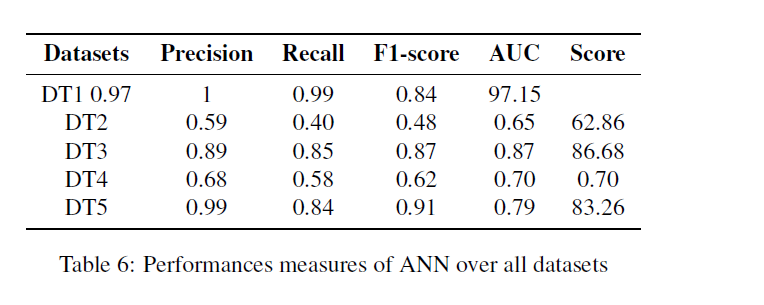
Table 5 shows the performance measures of the SVM classifier.

Table 1 shows the performance measures. The observation shows that the best score, precision and F1-score are obtained on the datasets DT1, DT3 and DT5. However the higher AUC and the best specificity are observer on the datasets DT1, DT3 and DT4.



**ANN**

The performance of the ANN varied throughout the study depending on the dataset, although overall it performed well. A large amount of initial effort was required to train and validate the model. Notice that best precision are achieved by ANN classifier on the datasets DT1, DT3 and DT5 which are respectively 97%, 89% and 99%. While the higher AUC and the best scores are obtained on the datasets DT1 and DT3. The Sensivity values are higher than specificity values on datasets



**Discussion**

Dans cette étude, les algorithmes DT, RF, LR, NB, SVM et ANN ont été appliqués sur cinq ensembles de données concernant des patients atteint ou non de paludisme et vivant dans des regions du Senegal a savoir: Diourbel, Thies et Fatick. En effet, dans l’optique de proposer une nouvelle techniques de diagnostic et de prédiction du paludisme, il est important de connaitre les performances de ceux existant à travers nos jeux de données. Une analyse minutieuse des performances de nos six classifieurs sur les cinq jeux de données a montré qu'il n'y a pas nécessairement un seul meilleur algorithme de classification, mais que l'algorithme le plus performant dépendra des caractéristiques de l'ensemble de données à analyser. Ainsi on remarque que tous les algorithmes produisent leurs meilleures précisions sur les jeux des données DT1, DT3, et DT5. Ces valeurs qui atteignent par moment 97% dépassement assez largement le test de diagnostic rapide largement utiliser dans les structure de sante notamment au Senegal. Cependant sur ces mêmes jeux de données les algorithmes présentent des spécificités souvent très faible par exemple 0.05 sur DT1. Ceci montre que nos classifiers les plus performants sont uniquement performant pour prédire une seule classe: soit le patient est atteint de paludisme soit il ne l’est pas mais pas dans les deux taches. Cela s’explique par le fait que les jeux des données DT1 et DT3 sont très déséquilibres. En fait dans ces jeux de données soit le nombre patient atteint de paludisme est plus grand que ceux qui ne l’ont pas soit c’est le contraire. Par ailleurs on note que sur les jeux donnés DT2 et DT4 tous les algorithmes présentent des specificités et des sensivités significatives et asses similaires. Contrairement à ce qui est cité un peu en haut, sur ces jeux de données les algorithmes sont performants sur les taches de prediction des deux classes. En regardant de près les résultats en termes de précision, de rappel et de F-mesure on observe que les classifieurs RF, LR, SVM et ANN surpassent les autres en général pour chaque jeux de donnés. En effet pour le de données DT1, qui contient des observations sur des patients vivant dans différentes régions du Sénégal, ces quatre classificateurs ont une précision de 99%, un rappel supérieur à 92% et une F-mesure supérieure à 95%. On note la même tendance avec le jeu de données DT2 qui contient des observations sur des patients vivant dans la même zone au Sénégal. On peut également remarquer que RF, LR, SVM et ANN présentent une meilleure précision que le test de diagnostic rapide réalisé et utilisé systématiquement au sein de la majorité des structures de santé au Sénégal. Ce constant reste vrai avec DT4 qui est un jeu de données parfaitement équilibré. En somme il nous est très difficile voire impossible de dire de manière définitif quel algorithme est plus performant pour la tache de prédiction du paludisme mais que le choix de celui-ci dépendra fortement du choix du jeu de données. Cependant cette étude montre une prise en charge de notre problème de classification une méthode intégrant plusieurs models et des jeux de données diverses s’avère nécessaire

La recherche a montré avec un accent particulier sur les données sur les soins de santé, qui sont discutées dans l'article., les résultats de la section précédente plaident clairement en faveur des classificateurs RF, LR, SVM avec noyau gaussien et ANN. En effet, compte tenu de la base de données DT1, qui contient des observations sur des patients vivant dans différentes régions du Sénégal les patients, ces quatre classificateurs ont une précision de 99%, un rappel supérieur à 92% et un

Mesure F supérieure à 95%. On note la même tendance avec le jeu de données DT2 qui contient des observations sur des patients vivant dans la même zone au Sénégal. Donc, en termes de précision, de rappel et de mesure F quatre classificateurs surpassent les autres. On peut également remarquer que RF, LR, SVM avec noyau gaussien et ANN présentent une meilleure précision que le test de diagnostic rapide réalisé et utilisé systématiquement au sein de la majorité des structures de santé au Sénégal. La différence des performances de nos classificateurs sur les deux ensembles de données s’explique par le fait que des facteurs climatiques tels que la température et les eaux stagnantes sont très déterminantes dans l'apparence ou non du paludisme dans des zones distinctes du Sénégal car ils favorisent le développement des moustiques responsables de la maladie. Lorsque nous nous limitons maintenant aux courbes ROC, nous observons que SVM avec noyau gaussien et Naive Bayes présente le meilleur taux de prédiction positive car ils ont les valeurs AUC les plus basses par rapport aux autres classificateurs. De plus, nous avons essayé d'étudier l'impact de la profondeur des arbres, le nombre de couches cachées et le nombre d'estimateurs pour DT, ANN et RF respectivement. Alors que l'augmentation de la profondeur de l'arbre diminue la précision de DT (voir figure 6), la plus haute précision de RF correspond à l'utilisation de 50 estimateurs comme le montre la figure 7. De plus, la figure 8 montre que lorsque le nombre de couches cachées augmente la précision de ANN le fait en général. Pour conclure, nous peut affirmer que ANN semble être la classification la plus prometteuse approche parmi les six algorithmes ML étudiés lorsque nous ne sommes intéressé par la précision, le rappel et la mesure F. Cependant si nous inclure les courbes ROC dans l'analyse SVM avec noyau gaussien devenir l'approche la plus efficace

The research has shown that there is not necessarily a single best classification tool, but instead the best performing algorithm will depend on the features of the dataset to be analysed, with particular emphasis on health care data, which are discussed in the paper.

**Conclusion**

This research has indicated that in practice there is no single best classification tool, but instead the best technique will depend on the features of the dataset to be analysed and any preferences of end-users. The research has made a start in investigating what these features are with particular emphasis on health care data. A summary of the main findings are as follows: