I.J. Information Technology and Computer Science, **2021, 6, 61-71**
Published Online December 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2021.06.05

Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation

Isaac Kofi Nti
Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana
E-mail: ntious1@gmail.com

Owusu Nyarko-Boateng
Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana
E-mail: owusu.nyarko-boateng@uenr.edu.gh

Justice Aning
Department of Computer Science, Sunyani Technical University, Sunyani, Ghana
E-mail: aning421@gmail.com

Received: 14 June 2021; Revised: 11 August 2021; Accepted:26 September 2021; Published: 08 December 2021

Abstract: The numerical value of k in a k-fold cross-validation training technique of machine learning predictive models is an essential element that impacts the model's performance. A right choice of k results in better accuracy, while a poorly chosen value for k might affect the model's performance. In literature, the most commonly used values of k are five (5) or ten (10), as these two values are believed to give test error rate estimates that suffer neither from extremely high bias nor very high variance. However, there is no formal rule. To the best of our knowledge, few experimental studies attempted to investigate the effect of diverse k values in training different machine learning models. This paper empirically analyses the prevalence and effect of distinct k values (3, 5, 7, 10, 15 and 20) on the validation performance of four well-known machine learning algorithms (Gradient Boosting Machine (GBM), Logistic Regression (LR), Decision Tree (DT) and K-Nearest Neighbours (KNN)). It was observed that the value of k and model validation performance differ from one machine-learning algorithm to another for the same classification task. However, our empirical suggest that k = 7 offers a slight increase in validations accuracy and area under the curve measure with lesser computational complexity than k = 10 across most MLA. We discuss in detail the study outcomes and outline some guidelines for beginners in the machine learning field in selecting the best k value and machine learning algorithm for a given task.

Index Terms: Cross-Validation, K-Fold, Leave-one-out, Machine learning, Computational complexity.

1. Introduction

Artificial intelligence (AI) is a discipline that seeks to enhance computers and software with human abilities to think and act like a human effectively. Since its inception, AI has brought tremendous worldwide economic growth in every sector of the economy and has augmented humanity's activities, helping us to make informed decisions [1–3]. AI has made everyday life safer, convenient and more straightforward, e.g., better monitoring and diagnostic systems in the healthcare industry for professionals to make better decisions and 24/7 availability of AI-enabled customer-centred telemedical services. In addition, AI-enabled self-driving vehicles are operating on a mobile ad-hoc network (MANET), enabling vehicle-to-vehicle (V2V) communications in the automotive industry. Several subfields of AI, such as machine learning (ML) and deep learning (DL), make it possible for a machine to obtain high-level human intelligence as applied in the healthcare and the automobile industry. Fig. 1 shows the various sub-fields of AI. In machine learning, one technique used for accessing training models that have received much attention over the years is cross-validation (CV), also known as rotation estimation or out-of-sample testing technique [4,5]. CV is typically used where ML task is prediction, and one seeks to evaluate how accurately a predictive model will perform in training [6]. An ML model is given a labelled dataset call training data (DTrain) on which model training is carried out in a classification or prediction task. The trained model is then tested on an unseen dataset called a testing set or validation dataset (D_{valid}).

The goal of a CV is to evaluate the model's ability to classify new data that was not employed in estimating the model, to flag problems like selection bias or overfitting [7]. In addition, for offering a piece of in-depth knowledge into how a trained model will generalise to an unseen dataset or a real-world problem.

There are two types of CV: exhaustive cross-validation (ECV) and non-exhaustive cross-validation (NECV). ECV methods are CV techniques that learn and test every possible way to split the original dataset into a D_{Train} and a D_{valid}. E.g., (i) leave-p-out CV and (ii) leave-one-out CV (LOOCV). One key drawback of the ECV techniques is the computational complexity. NECV techniques do not compute every possible way of dividing the original dataset, approximating leave-p-out CV. E.g., (i) k-fold CV, (ii) holdout CV and (iii) repeated random sub-sampling CV. Of all the available CV techniques, one commonly used among machine learning practitioners in the k-fold CV [8], due to its user-friendliness and powerful nature in evaluating the success rate of ML models used for classification [5,6,9]. However, one critical issue with the k-fold CV is selecting the best value of k.

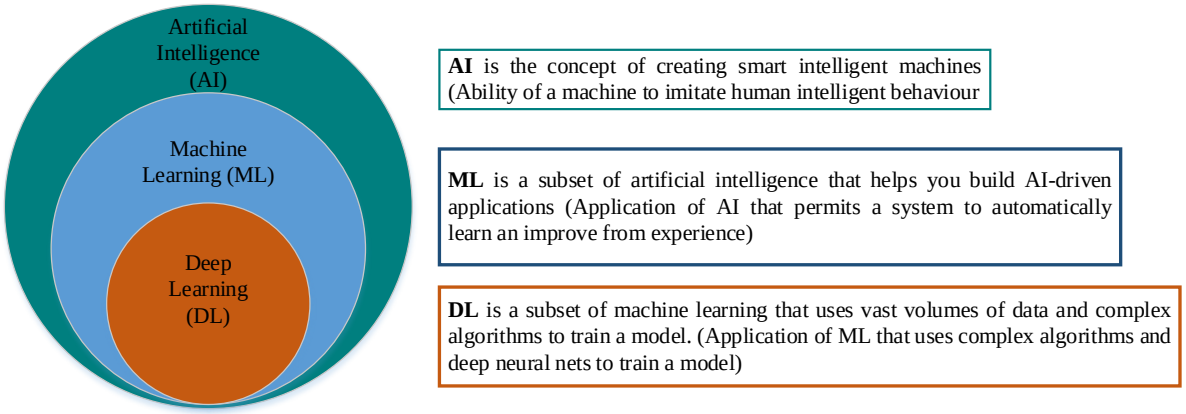

Fig.1. Artificial Intelligence and various subfields

Hence, we focus on the k-fold CV's problem in this study to examine the effect of different k values on the validation performance of four well-known machine learning algorithms. However, the core questions guiding this study remains:

1. What is the effect of different k values in k-fold cross-validation on various machine learning algorithms?
2. What is the optimal value of k that improves the validation performance of most machine learning algorithms?

The remaining sections of this paper are grouped as follows: Section 2 presents a summary of related studies, the methodology of the study is present in section 3. Finally, we present the results and their implications in section 4, while the concludes the study and directions for future works are presented in section 5.

2. Related Works

The critical nature of k values in a k-fold cross-validation training in machine learning applications has resulted in several researchers looking at the possible optimal k for most machine learning algorithms. This section presents a few of these studies.

The best k value for k-fold CV training for KNN, decision trees, Naïve Bayes, SVM, Linear Discriminate Analysis (LDA) and Linear Regression CART (Classification and Regression Tree) using similar characteristics on crime data was undertaken [9]. The study outcome showed that KNN outperformed all other classifiers trained with the same k value. Similarly, the validation accuracy of four different k values (k = 3, 5, 7, and 10) for training Single ANN (SANN) and Ensemble ANN (EANN) were examined [6]. The results show that the use of k = 10 offers better performances and less biased estimates. Likewise, different k values (3, 5, 7, and 9) was used to train the KNN algorithm for image classification. The study concludes that k = 3 gives better classification accuracy than 5, 7 and 9. Marcot and Hanea [5] adopted different k values ((k = 2, 5, 10, 20, n − 5, n − 2, and n − 1)) in a k-fold CV for training discrete Bayesian network. The study outcome supported the common use of k = 10 in the literature; however, in some cases, k = 5 offered better accuracy with Bayesian network models. Our partial search of the literature shows a high percentage of ML practitioners use k = 10 [10–19] and k = 5 [9,20,21]. Only a few studies used k = 2 or 3 [22,23]. The literature suggests that k =10 is widespread among ML professionals, but as argued in the literature, there is no formal rule [5,20,22,24]. Nevertheless, few studies have comprehensively explored how different values of k apart affects the validation results in different ML models tested with a dataset of known statistical properties.

3. Methodology

This section presents the methods and tools used to achieve the aim of this study. Firstly, a framework is designed, we download the study dataset, prepare it. We then selected our machine learning algorithms and trained them on our dataset with different values of k values in k-fold CV. Next, we measured the validation accuracy by each classifier for each k value using two well-known evaluators. Finally, a comparative analysis of the results is carried out.

3.1. Research Design

Fig. 2 shows the workflow diagram of the current study, and it has five (5) stages. **Stage 1 (Dataset)**: The dataset used in this paper is open-source data downloaded from the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets.html) of size (rows = 5110, columns = 12). The dataset is an observation on stroke-related incidence used to predict whether a patient is likely to get a stroke using independent features such as various diseases, gender, smoking status, and age. **Stage 2 (Data Preprocessing)**: We preprocessed the dataset to remove errors, redundant data, contradictions and blank spaces. Finally, mean values were used to replace missing and NAN's values in our dataset. The computational complexity efficiency of machine learning models depends on the scale of features; a uniform feature promotes better prediction efficiency. Hence, our dataset was scaled between 0 and 1, as defined in Eq. (1).

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

$$\mu = \frac{sum(x)}{count(x)} \tag{2}$$

$$\sigma = \sqrt{\left\{\frac{sum((x-\mu)^2)}{count(x)}\right\}} \tag{3}$$

where x' is the scaled feature, x the feature to be scaled, μ is the mean and σ is the standard deviation

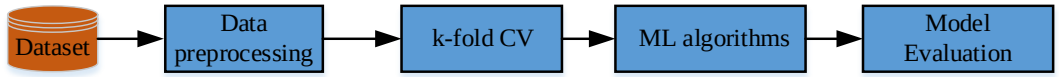

Fig.2. Study workflow diagram

Stage 3 (K-Fold Cross-Validation): Fig. 3 shows an overview of *k-fold* CV. A *k-fold* CV is a technique used for evaluating MLM. In k-fold CV, the dataset is divided into k fold. A fold is used in each iteration once as testing data, while the remaining folds are used as training data [24]. Thus, the process is repetitive until all dataset is evaluated. *K-Fold* CV results are normally re-iterated with the mean score of the values of the MLM. In this study, different k values [3, 5, 7, 10, 15 and 20] were selected and compared to identify the optimal k value that improves accuracy in a classification task.

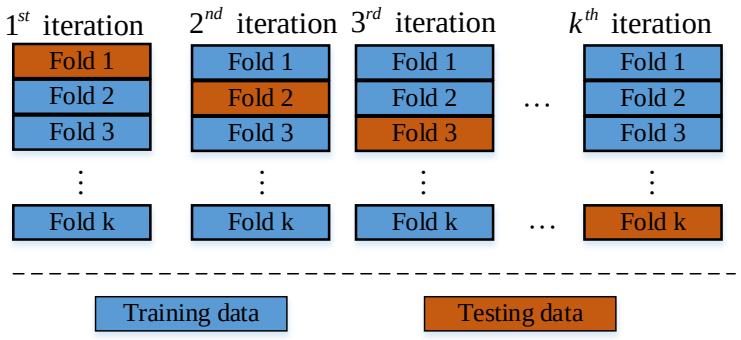

Fig.3. K-fold CV

Steps for k-fold CV

*Step*1 → devide the training dataset (D_{train}) into k equal subsets such as $\{f_1, f_2, \cdots, f_k\}$

*Step*2 → *for* $i = 1$ *to* $i = k$

*Step*3 → Let k -1 folds be the CV training dataset and f as the testing dataset

Step4 → using CV train the machine learning model (m) and compute the accuracy (Acc)

Step5 → Evaluate the Acc using all the k instances of CV

Stage 4 (Machine Learning Model (MLM)): An MLM is a mathematical representation of a real-world scenario. In this paper, four (4) well-known machine learning algorithms were adopted for training and prediction based on different k-fold CV.

1. **Gradient Boosting Machine (GBM)** put together the estimates from multiple DTs to make the final predictions. It is one of the most potent MLA building predictive both regression and classification models. GBM's ensemble shallows DTs sequentially with each DT learning and improving on the earlier one, whereas RF builds an ensemble of deep autonomous DTs [25].
2. **Logistic Regression (LR)** is an MLA describing data and explaining the association between one dependent binary feature and two or more ordinal, nominal, interval or ratio-level independent features [26].
3. **Decision Tree (DT)** is a simple supervised MLA for both regression and classification tasks; however, it is mainly used to classify problems. It is tree-like in construction, where inner nodes characterise the features of a dataset, branches signify the decision rules, and each leaf node signifies the result [10].
4. **K-Nearest Neighbours (KNN)** is easy to understand and implement supervised MLA for classification or regression tasks. It adopts the resemblance between new data and available data and puts the new data into the most similar group to the available groups. It is a non-parametric algorithm, i.e., it does not make any guess on primary data. It is occasionally referred to as a lazy learner algorithm since it does not learn from the training dataset instantly instead keeps it. At the classification period, it completes an action on the dataset.

We trained each classifier with all possible k values and recorded its classification performance based on two evaluators. Next, we repeat the process until all classifiers are trained with each possible k value, as shown in Algorithm 1.

Algorithm 1

```
read(k)                    {k value as a vector}
read (D_train)             {training dataset}
read classifiers(cl)       {cl selected classifier as a list}
for cl_i in cl             {for each classifier in cl}
→ for k_i of k             {for each value of k}
      devide (D_train) into k_i
      using CV train (cl_i) and calculate the accuracy
      calculate the performance of (cl_i) with all the k_i instances of CV
→ end for
end for
result(cl_i)
```

Stage 5 (Model Evaluation): We evaluated the k-fold CV training performance of the selected ML algorithms based on two well-known metrics for evaluating classification tasks, namely accuracy (Acc) and area under the curve (AUC), as defined in Eq. (4) and (5).

$$Accuracy(\%) = \left(\frac{(TP + TN)}{(TP + FN + FP + TN)} \right) \times 100 \tag{4}$$

$$AUC = \int_0^1 \frac{TP}{(TP + FN)} d \frac{FP}{(FP + TN)} = \int_0^1 \frac{TP}{P} d \frac{FP}{N} \tag{5}$$

where, TP= true positive, TN = true negative, FN = false negative, FP = false positive

3.2. Experimental Setup

For generalisation purposes, feature selection or dimensionality reduction was not carried, i.e., we strictly considered the original features for training and testing. The predictive performance of MLAs is generally affected by assigned hyperparameters values. Taking this into thought, in our try-outs, we used the default scikit-learn hyperparameters (see Table 1). All empirical analysis in this paper was done using the Python programming language with the Pandas, Scikit-Learn, and Matplotlib. A Lenovo (20EGS12E00), Intel® core™ i5-4340M CPU @ 2.90GHz (4 CPUs) 12GB memory was used in this study.

Table 1. Hyperparameter Settings

Models	Hyperparamters
GBM	loss='deviance', learning_rate=0.1, n_estimators=100
LR	penalty= l2, solver='lbfgs', max_iter =100, multi_class='auto', tol=0.0001, C=1.0
DT	criterion='gini', splitter='best', max_depth=None, min_samples_split=2,
KNN	n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30, p=2

4. Results and Discussions

4.1. Results

The core question for the current study is to examine if there is an optimal value of k, which enhances MLM validation performance. We examined different k values [3, 5, 7, 10, 15 and 20] on the cross-validation performance of four different machine learning algorithms to predict the likelihood of someone getting a stroke. A minimum k-value without undue computational complexity that achieved the best CV performance in prediction accuracy (Acc) and AUC is taken as the optimal k-value. Table 2 shows a comparative mean summary for all values of k in this study. We observed that a specific value of k, say *x*, results in different performance among machine-learning algorithms for the same classification task. As seen in Table 1, the accuracy of the logistic regression models is constant (0.959) irrespective of the variation in the value of k. However, we observed slightly decreases in the AUC measure of the LR, as '*k*' increased from 3 to 15. From the outcome, it suggests that k =3 is more suitable for the LR classifier in predicting stroke. The accuracy and AUC measure GBM and KNN increased steadily for k values of 3 and 7. The GBM and KNN recorded the optimal accuracy at k = 7, while the DT obtained the best accuracy at k = 15. Thus, the outcome suggests that the validation performance of a machine learning algorithm is partly dependent on the value of k.

Table 2. Comparative summary of mean results for all k-values

	GBM		LR		DT		KNN	
k	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**
3	0.951	0.8	**0.959**	**0.832**	0.922	0.538	0.956	0.547
5	0.954	0.801	0.959	0.818	0.923	0.52	0.954	0.589
7	**0.955**	0.822	0.959	0.83	**0.929**	0.561	**0.957**	0.633
10	0.953	0.813	0.959	0.828	0.924	0.559	0.956	0.621
15	0.951	**0.823**	0.959	0.831	**0.93**	**0.581**	0.955	**0.638**
20	0.949	0.82	0.959	0.832	0.927	0.547	0.956	0.649

Table 3. Model training performance with k =3

	GBM		LR		DT		KNN	
k	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**
1	0.956	0.815	0.958	0.845	0.933	0.543	0.953	0.53
2	0.948	0.749	0.96	0.811	0.913	0.505	0.96	0.587
3	0.95	0.836	0.958	0.838	0.921	0.565	0.955	0.525
Mean	0.951	0.8	**0.959**	**0.832**	0.922	0.538	0.956	0.547
SD	0.03	0.037	0.001	0.015	0.008	0.025	0.003	0.028

Table 3-6 shows the MLA performance for k = 3, 5, 7 and 10, respectively. For a k value of 3, the LR outperformed the GBM, DT and KNN. On the other hand, it was observed that the classification accuracy and AUC for GMB, DT and KNN decreased with increasing values of k, again more or less stabilising with k ≥ 10 to 20. The results (see Table 3 -10) suggest that the consequence of increasing the value of k in a k-fold CV on classification accuracy and AUC results in an initial surge in classification accuracy with low values of k-folds (say 3, 5, 7 and 10), however,

decreases progressively with greater values of k (15 and 20). The mixed performance results of the machine learning algorithms (MLA) with different k values suggest k cannot be generalised for all MLA. Thus, the outcome suggests the performance of an MLM when subjected to different k-fold CV is likely determined by different factors relating to the model architecture and complexity and the degree and multiplicity of the applied dataset. Tables 3-6 shows that k = 7 is the optimal k-value as it offered the best accuracy and AUC measure for GBM, LR, DT and KNN machine learning models.

Table 4. Model training performance with k =5

k	GBM		LR		DT		KNN	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
1	0.951	0.792	0.959	0.83	0.914	0.524	0.947	0.592
2	0.959	0.811	0.959	0.823	0.926	0.483	0.951	0.633
3	0.947	0.737	0.959	0.759	0.934	0.535	0.959	0.626
4	0.955	0.881	0.959	0.868	0.93	0.533	0.959	0.581
5	0.959	0.782	0.959	0.81	0.913	0.524	0.955	0.512
Mean	**0.954**	**0.801**	**0.959**	**0.818**	**0.923**	**0.52**	**0.954**	**0.589**
SD	0.005	0.047	0	0.035	0.009	0.019	0.005	0.043

Table 5. Model training performance with k =7

k	GBM		LR		DT		KNN	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
1	0.96	0.828	0.96	0.833	0.937	0.556	0.96	0.539
2	0.954	0.828	0.954	0.852	0.925	0.544	0.954	0.547
3	0.965	0.724	0.96	0.822	0.919	0.547	0.954	0.796
4	0.948	0.816	0.96	0.752	0.948	0.562	0.96	0.568
5	0.954	0.836	0.96	0.83	0.942	0.559	0.96	0.657
6	0.954	0.862	0.96	0.909	0.908	0.473	0.954	0.64
7	0.954	0.858	0.96	0.81	0.925	0.687	0.96	0.686
Mean	**0.955**	**0.822**	**0.959**	**0.83**	**0.929**	**0.561**	**0.957**	**0.633**
SD	0.005	0.043	0.002	0.044	0.013	0.059	0.003	0.085

Table 6. Model training performance with k =10

k	GBM		LR		DT		KNN	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
1	0.959	0.838	0.959	0.802	0.934	0.583	0.959	0.591
2	0.943	0.802	0.959	0.822	0.893	0.466	0.959	0.673
3	0.959	0.708	0.959	0.872	0.934	0.583	0.959	0.449
4	0.959	0.807	0.959	0.843	0.934	0.678	0.942	0.84
5	0.934	0.686	0.959	0.817	0.917	0.67	0.959	0.601
6	0.95	0.743	0.959	0.7	0.909	0.474	0.959	0.643
7	0.959	0.881	0.959	0.879	0.942	0.587	0.959	0.547
78	0.959	0.919	0.959	0.872	0.917	0.478	0.959	0.647
9	0.959	0.838	0.959	0.853	0.934	0.487	0.95	0.616
10	0.95	0.907	0.959	0.816	0.926	0.578	0.959	0.601
Mean	**0.953**	**0.813**	**0.959**	**0.828**	**0.924**	**0.559**	**0.956**	**0.621**
SD	0.008	0.076	0	0.05	0.014	0.075	0.005	0.094

To assess the best k values (7 and 10) from our experiments, we trained the selected MLAs with the leave-one-out cross-validation (LOOCV) technique and compared their outcomes to our optimal k values. We then calculated the correlation between the average classification accuracy from the 7-fold CV, 10-fold CV and the LOOCV. Fig. 6 compares models' accuracy with k-fold (k = 5 and k = 10) and LOOCV. We observed that for some algorithms, the accuracy by the k-fold is infinitesimal higher than that obtained with the LOOCV, and in other cases, its under-estimates the accuracy. Fig. 7 shows a scatter plot comparing the distribution of mean accuracy scores of 10-fold (x-axis) vs the LOOCV (y-axis). Fig. 8 shows a scatter plot comparing the distribution of mean accuracy scores of 7-fold (x-axis) vs the LOOCV (y-axis). We observed a high linear positive correlation (0.999) between 10-fold CV and LOOCV. In contrast, a perfect correlation of (1.000) between 7-fold and LOOCV. The results support the earlier discussion that k = 7 leads to a better MLA model performance than k = 10.

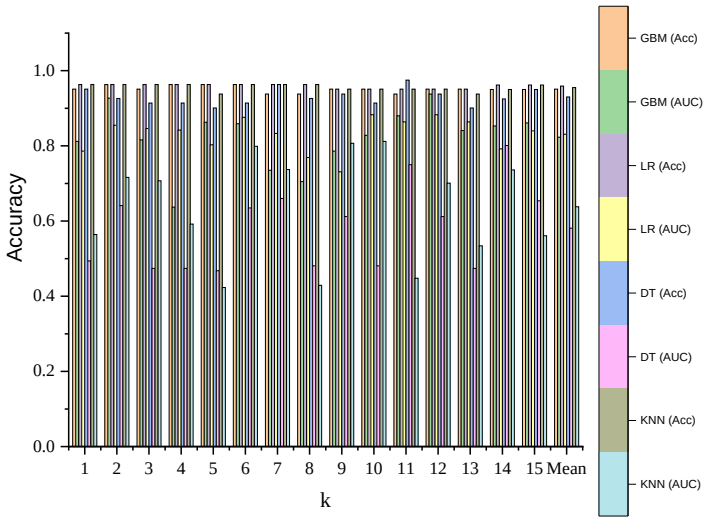

Fig.4. Model training performance with k =15

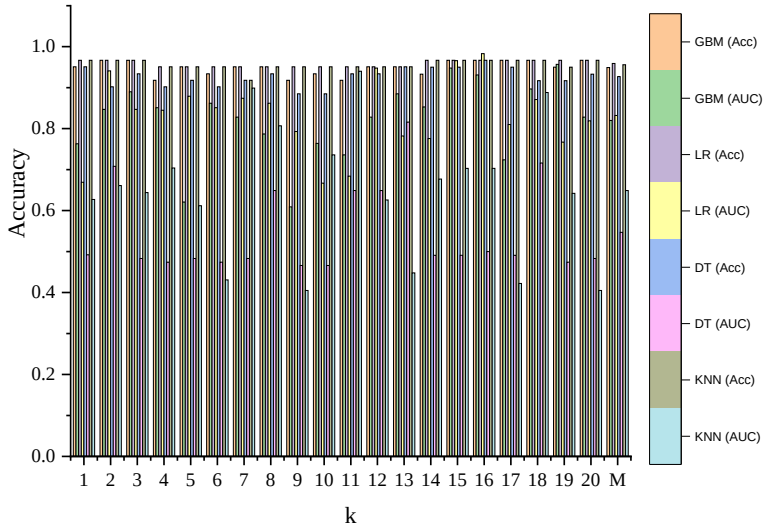

Fig.5. Model training performance with k =20

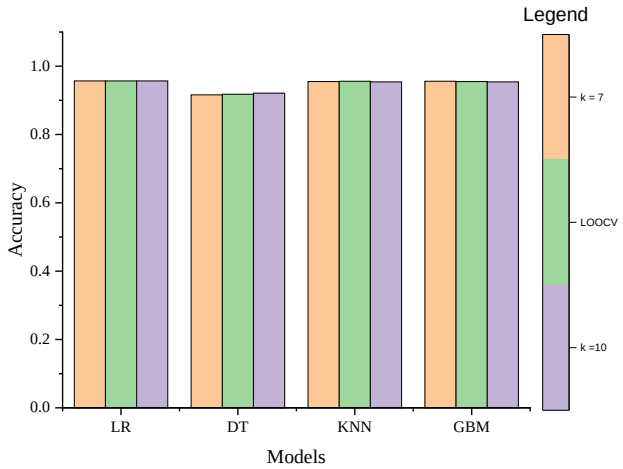

Fig.6. Comparison of models' performance on k = 5 and k = 10 with LOOCV

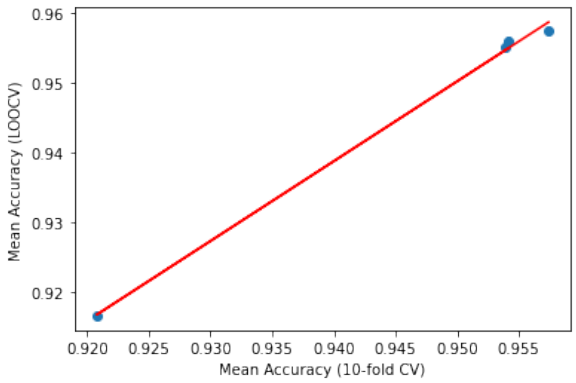

Fig.7. Distribution of mean accuracy scores k-fold (k = 10) and LOOCV

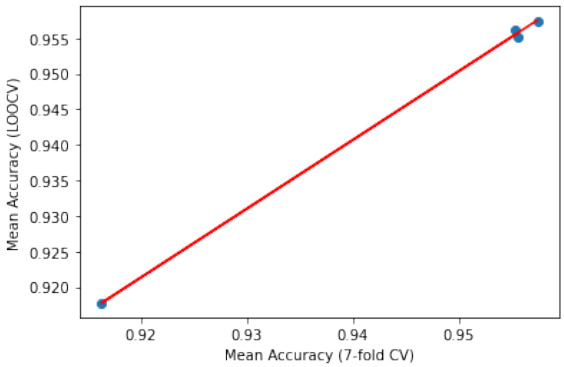

Fig.8. Distribution of mean accuracy scores k-fold (k = 7) and LOOCV

4.2. Discussion

The essential configuration parameter for k-fold CV is *k*, which describes the number of folds in which a given dataset for a machine learning task is partitioned. This study has performed a comparative study of different k values in k-fold CV on Gradient Boosting Machine (GBM), Logistic Regression (LR), Decision Tree (DT) and K-Nearest Neighbours (KNN) classification algorithms. The aim was to establish an optimal k value, which gives better prediction accuracy and AUC. In this study, the optimal k value was defined as the smallest number of folds (to prevent unnecessary computational complexity) which the classification error rate of a classifier is steady.

Several unanticipated associations arose from our investigation. The experimental results above show that LR accuracy and AUC measures are independent of the variation in k value. However, the GBM and KNN vary as the value of k increases from 3 to 10 and drops, as '*k*' is >10. The DT obtained high prediction accuracy as k =15. The difference in prediction accuracies with variation in k among the MLAs suggests that the optimal value of k differs among MLAs. Our findings disagree with [5] that recorded k =10 as the best k values for a classification task using Bayesian network and [6] that obtained k = 10 as more suitable than k = 3, 5 and 7 for nitrate load estimations using neural networks. Likewise, in [24], it was observed that k = 3 offered the best classification accuracy with KNN. The different optimal *k* in literature based on distinct ML algorithms shows that *k* is partial ML algorithm dependent. Fig. 4 and 5 show the models performance for k = 15 and k =20, respectively.

The results suggest that the k-Fold CV algorithm can enhance prediction accuracy and get the best machine learning algorithm for build a more robust predictive model. The results suggest that perhaps k = 5 and k = 10 alone are slightly optimistic, and possibly k = 7 is a more accurate estimate. The study affirms literature [5,6,24] that the higher the value of k, the more the computational complexity; almost linearly (asymptotically, linearly) for training MLAs with algorithmic complexity linear in the total quantity of training examples. The computational time for training increases *k*-1 times if the training time is linear in the number of training instances. So, we recommend considering accuracy and generalisation for a smaller dataset, especially since one needs to get the best out of inadequate training instances. However, for extensive training datasets and MLAs with high asymptotical computational complexity growth in the number of training instances (at least linear), we suggest *k* = 3-5. Thereby preventing the possibility of excessive increase in computational time for a training MLA with asymptotic complexity linear in the number of training samples.

The accuracies recorded by the MLAs with LOOCV were seen to be a little higher than the k-fold (k = 7 and k = 10) (see Fig. 6), making the LOOCV an ideal technique for training and comparing the performance of MLAs. However, the run-time of the LOOCV compared with k-fold (in this case, k =7 and k =10) was significantly higher. Thus, the k-fold is far computational efficient than the LOOCV.

5. Conclusions

Cross-validation (CV) is one of the most used techniques in measuring the optimal model in machine learning. Of the various type of CV, the k-fold is common among machine learning practitioners. However, randomly selecting the k value can lead to poor machine learning algorithm performance and computational complexity. Also, literature has shown that the value of k and model validation performance differs among machine learning algorithms. Thus, it is vital to select an optimal value of k in k-fold CV. However, to the best of our knowledge, a few if any studies have explored in detail with extensive empirical results how values of k (number of subsets) affect validation results off different machine learning algorithms.

Therefore, this paper investigates the effect of different k values (3, 5, 7, 10, 15 and 20) in k-fold CV on machine learning algorithm performance. We trained gradient boosting machine, logistic regression, decision tree and K-nearest neighbours classifiers with each k value to predict whether a patient is likely to get a stroke. We observed that k and model validation performance values differ from one machine-learning algorithm to another for the same classification task. Most uses of k = 10 in the literature (as discussed above) can be supported by our results. However, in many cases, k = 7 offers a slight difference in accuracy and area under the curve measure with lesser computational complexity than k = 5 or 10 for all selected machine learning algorithms (see Table 5). A correlation coefficient of 1.000 achieved between the mean accuracy by models for k = 7 and LOOCV compared of 0.999 for $k = 10$ affirms the optimality of k = 7. The outcome reveals that we cannot always expect to have a more accurate result by increasing or decreasing folds. In some instances, the increase in the value of k will improve the accuracy, and occasionally, it will only increase the computational complexity.

Similarly, there is no direct link between increasing or decreasing the value of K in k-fold CV and increasing or decreasing the accuracy of the machine learning algorithms. Therefore, in selecting the value of k, one needs to be cautious since a lower k value is less expensive in terms of computational complexity, less variance, however, more bias. In contrast, a higher value of k is computationally expensive but has more variance and lower bias. Thus, the value of k must permit the size of each validation set to be adequate to warrant a fair estimate of the model's performance. Simultaneously k should not be tiny, e.g., two such that trained models are not scarce to evaluate. So, it is of essence to perform experiments with different k values for a given dataset and algorithm to find the optimal k value for accuracy improvement.

Acknowledgement

We want to express our appreciation to all those who have supported us during our research in the Department of Computer Science and Informatics at UENR and Computer Science at STU, Sunyani.

References

[1] Dayana C. Tejera Hernández,"An Experimental Study of K* Algorithm", International Journal of Information Engineering and Electronic Business, vol.7, no.2, pp.14-19, 2015.

[2] Shaimaa Mahmoud, Mahmoud Hussein, Arabi Keshk, "Predicting Future Products Rate using Machine Learning Algorithms", International Journal of Intelligent Systems and Applications, Vol.12, No.5, pp.41-51, 2020.

[3] Seyyid Ahmed Medjahed, Mohammed Ouali, Tamazouzt Ait Saadi, Abdelkader Benyettou,"An Optimization-Based Framework for Feature Selection and Parameters Determination of SVMs", International Journal of Information Technology and Computer Science, vol.7, no.5, pp.1-9, 2015.

[4] D.M. Allen, The Relationship between Variable Selection and Data Agumentation and a Method for Prediction, Technometrics. 16 (1974) 125. https://doi.org/10.2307/1267500.

[5] B.G. Marcot, A.M. Hanea, What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?, Comput. Stat. (2020). https://doi.org/10.1007/s00180-020-00999-9.

[6] K. Jung, D.H. Bae, M.J. Um, S. Kim, S. Jeon, D. Park, Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with K-fold cross-validation, Sustain. 12 (2020). https://doi.org/10.3390/SU12010400.

[7] C.R. Rao, Y. Wu, Linear model selection by cross-validation, J. Stat. Plan. Inference. 128 (2005) 231–240. https://doi.org/10.1016/j.jspi.2003.10.004.

[8] S. Geisser, The Predictive Sample Reuse Method with Applications, J. Am. Stat. Assoc. 70 (1975) 320–328. https://doi.org/10.1080/01621459.1975.10479865.

[9] P. Tamilarasi, R.U.U. Rani, Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning, in: 2020 Fourth Int. Conf. Comput. Methodol. Commun., IEEE, 2020: pp. 1034–1038. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000193.

[10] I.K. Nti, A.F. Adekoya, B.A. Weyori, A comprehensive evaluation of ensemble learning for stock-market prediction, J. Big Data. 7 (2020) 20. https://doi.org/10.1186/s40537-020-00299-5.

[11] M. Barstugan, U. Ozkaya, S. Ozturk, Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods, (2020). http://arxiv.org/abs/2003.09424.

[12] S. Tuarob, C.S. Tucker, M. Salathe, N. Ram, An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages, J. Biomed. Inform. 49 (2014) 255–268. https://doi.org/10.1016/j.jbi.2014.03.005.
[13] B.A. Tama, S. Im, S. Lee, Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble, Biomed Res. Int. 2020 (2020) 1–10. https://doi.org/10.1155/2020/9816142.
[14] A. Oztekin, R. Kizilaslan, S. Freund, A. Iseri, A data analytic approach to forecasting daily stock returns in an emerging market, Eur. J. Oper. Res. 253 (2016) 697–710. https://doi.org/10.1016/j.ejor.2016.02.056.
[15] L. D, B. Vishnuvardhan, Classification Performance Improvement Using Random Subset Feature Selection Algorithm for Data Mining, Big Data Res. 12 (2018) 1–12. https://doi.org/10.1016/j.bdr.2018.02.007.
[16] K.T. Chui, R.W. Liu, M. Zhao, P.O. De Pablos, Predicting Students' Performance With School and Family Tutoring Using Generative Adversarial Network-Based Deep Support Vector Machine, IEEE Access. 8 (2020) 86745–86752. https://doi.org/10.1109/ACCESS.2020.2992869.
[17] S. Simsek, U. Kursuncu, E. Kibis, M. AnisAbdellatif, A. Dag, A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival, Expert Syst. Appl. 139 (2020). https://doi.org/10.1016/j.eswa.2019.112863.
[18] I.K. Nti, A.F. Adekoya, B.A. Weyori, Efficient Stock-Market Prediction Using Ensemble Support Vector Machine, Open Comput. Sci. 10 (2020) 153–163. https://doi.org/10.1515/comp-2020-0199.
[19] I.K. Nti, A.F. Adekoya, O. Nyarko-Boateng, A Multifactor Authentication Framework for the National Health Insurance Scheme in Ghana using Machine Learning, 13 (2020) 639–648. https://doi.org/10.3844/ajeassp.2020.639.648.
[20] O. Karal, Performance comparison of different kernel functions in SVM for different k value in k-fold cross-validation, in: 2020 Innov. Intell. Syst. Appl. Conf., IEEE, 2020: pp. 1–5. https://doi.org/10.1109/ASYU50717.2020.9259880.
[21] R. Ghorbani, R. Ghousi, Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques, IEEE Access. 8 (2020) 67899–67911. https://doi.org/10.1109/ACCESS.2020.2986809.
[22] L. Fang, S. Liu, Z. Huang, Uncertain Johnson–Schumacher growth model with imprecise observations and k-fold cross-validation test, Soft Comput. 24 (2020) 2715–2720. https://doi.org/10.1007/s00500-019-04090-4.
[23] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, J.R. Smith, Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images, in: L. Zhou, L. Wang, Q. Wang, Y. Shi (Eds.), Mach. Learn. Med. Imaging, Springer International Publishing, Cham, 2015: pp. 118–126.
[24] M.R. Wayahdi, D. Syahputra, S. Hafiz, N. Ginting, Evaluation of the K-Nearest Neighbor Model With K-Fold Cross Validation on Image Classification, J. Infokum. 9 (2020) 1–6.
[25] B. Boehmke, B.M. Greenwell, Hands-on machine learning with R, CRC Press, 2019.
[26] H. Ahmed, E.M.G. Younis, A. Hendawi, A.A. Ali, Heart disease identification from patients' social posts, machine learning solution on Spark, Futur. Gener. Comput. Syst. 111 (2020) 714–722. https://doi.org/10.1016/j.future.2019.09.056.

Authors' Profiles

Isaac Kofi Nti holds HND in Electrical & Electronic Engineering from Sunyani Technical University, B. Sc. in Computer Science from Catholic University College, M. Sc. in Information Technology from Kwame Nkrumah University of Science and Technology. Mr Nti is a Lecturer at the Department of Computer Science and Informatics, University of Energy and Natural Resources (UENR), Sunyani Ghana and currently is a Ph. D. candidate and the same department. His research interests include artificial intelligence, energy system modelling, intelligent information systems and social and sustainable computing, business analytics and data privacy and security.

ORCID ID: https://orcid.org/0000-0001-9257-4295

Owusu Nyarko-Boateng is a PhD Computer Science candidate and a lecturer at the Department of Computer Science and Informatics, the University of Energy and Natural Resources, Ghana. He holds HND Electrical & Electronics Engineering, BSc Computer Science, PGDE, and MSc Information Technology. Owusu Nyarko-Boateng has previously worked with MTN Ghana and Huawei Technologies (SA) for over ten (10) years. He has in-depth experience in telecommunications transmission systems, including fiber optics cable deployment for long-haul and short distance (FTTx), 2G BTS, WCDMA (3G), and 4G (LTE) plants installations and configurations. He also managed Huawei DWDM Optix8800 OSN, Huawei OptiX OSN1800 and optical distributor frames (ODF). As an academic and a researcher, Owusu Nyarko-Boateng has developed a passion in the following research areas: machine learning, artificial intelligence, computer networks and data communications, network security, fiber optics technologies, modelling transmission systems, 5G & 6G Technologies, Expert Systems, computational intelligence for data communications.

Justice Aning received the MSc. degree in Information Technology from The Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in 2017. He is currently a lecturer with the Department of Computer Science, Sunyani Technical University, Sunyani, Ghana. J. Aning has authored more than five papers in journals. His current research interests include Web design, Machine learning, intelligent systems for modelling and optimisation.

How to cite this paper: Isaac Kofi Nti, Owusu Nyarko-Boateng, Justice Aning, "Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation", International Journal of Information Technology and Computer Science(IJITCS), Vol.13, No.6, pp.61-71, 2021. DOI: 10.5815/ijitcs.2021.06.05