Scientific Research Publishing

Open Journal of Statistics, 2022, 12, 563-580
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

# Classification of Stateless People through a Robust Nonparametric Kernel Discriminant Function

**Macdonald G. Obudho, George O. Orwa, Romanus O. Otieno, Festus A. Were**

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
Email: mobudho@knbs.or.ke, gorwa@buc.ac.ke, rodhiambo@must.ac.ke, Weref87@gmail.com

**How to cite this paper:** Obudho, M.G., Orwa, G.O., Otieno, R.O. and Were, F.A. (2022) Classification of Stateless People through a Robust Nonparametric Kernel Discriminant Function. *Open Journal of Statistics*, **12**, 563-580.
https://doi.org/10.4236/ojs.2022.125034

**Received:** June 7, 2022
**Accepted:** October 8, 2022
**Published:** October 11, 2022



## Abstract

Statelessness is the absence of any Nationality. These include the Pemba, Shona, Galjeel, people of Burundi and Rwanda descent, and children born in Kenya to British Overseas Citizens after 1983. Frequently, they are not only undocumented but also often overlooked and not included in National Administrative Registers. Accordingly, find it hard to participate in Social and Economic Affairs. There has been a major push by UNHCR and international partners to "map" the size of stateless populations and their demographic profile, as well as causes, potential solutions and human rights situation. One of the requirements by the UNHCR in their push is for countries to find a potential solution to statelessness which starts with classifying/associating a person from these communities to a particular local community that is recognized in Kenya. This paper addresses this problem by adopting a Robust Nonparametric Kernel Discriminant function to correctly classify the stateless communities in Kenya and compare the performance of this method with the existing techniques through their classification rates. This is because Nonparametric functions have proven to be more robust and useful especially when there exists auxiliary information which can be used to increase precision. The findings from this paper indicate that Nonparametric discriminant classifiers provide a good classification method for classifying the stateless communities in Kenya. This is because they exhibit lower classification rates compared to the parametric methods such as Linear and Quadratic discriminant functions. In addition, the finding shows that based on certain similarities in characteristics that exist in these communities that surround the Pemba Community, the Pemba community can be classified as Giriama or Rabai in which they seem to have a strong link. In this regard, the study recommends the use of the Kernel discriminant classifiers in classifying the stateless persons and that the Government of Kenya consider integrating/recognizing

the Pemba community into Giriama or Rabai so that they can be issued with the National Identification Cards and be recognized as Kenyans.

## Keywords

Discriminant Analysis, Kernel Discriminant, Nonparametric, Classification, Statelessness

## 1. Introduction

Nationality acts as the linkage between a citizen and the international system through domestic laws. Nationality, traces its roots to the history of human race with human beings having a sense of belonging to a nation/country and hence the nationality to which an individual belongs guarantee him rights to citizen rights. Although, every person can have the right to nationality the same has not been experienced by every individual in the world. This has created a situation that has led to some individuals being stateless in their host country [1].

A stateless person is someone who, under National Laws, does not enjoy Citizenship—the legal bond between a government and an individual—in any country. Statelessness is a global anomaly and many persons who are stateless have never crossed an international border [2] [3]. Two United Nations Conventions established the international legal framework for the protection of stateless persons and the prevention and reduction of statelessness. The 1954 Convention Relating to the Status of Stateless Persons gives the definition of stateless persons and also provides important minimum standards of treatment for stateless persons. It defines a stateless person as a person who is not considered as a national by any state under the operation of its law. The 1961 Convention on the Reduction of Statelessness sets out guidelines for the prevention of statelessness.

Kenya has a few groups who remain in protracted statelessness situations. These include the Pemba, the Shona, people of Burundi and Rwanda descent, and children born in Kenya to British Overseas Citizens after 1983, [4]. These persons are not only undocumented but also often overlooked and not included in national administrative registers and databases. Many stateless persons and persons of undetermined nationality are counted in the defacto population and housing censuses but often go unrecognized by nationality or ethnic affiliation.

Although the number of stateless persons in Kenya is unclear, after the registration of the Makonde, an estimate of 18,500 stateless persons in Kenya is being used, [5]. Despite various amendments to provisions providing for the right to a nationality, many of Kenya's domestic laws on nationality are discriminatory and infringe greatly on the fundamental human rights, hence potentially resulting in an increase in the number of people that become stateless or those who are stateless remain that way indefinitely. Kenya has to date not ratified the 1954 Convention relating to the Status of Stateless Persons and the 1961 Convention

on the Reduction of Statelessness. Nevertheless, the discriminatory nationality laws and the administration thereof have repeatedly been brought to the attention of the international human rights community. The grounds thereof are based on Kenya's national laws being inconsistent with Kenya's international human rights obligations. In order to make an adequate assessment of Kenya's national laws, it should be noted that the causes of statelessness in Kenya can be divided into two broad categories, namely, administrative and legal, which illustrates the gap between law and practice.

The administrative causes of statelessness in Kenya such as the faulty operation or under-regulated nature of Kenya's administrative practices concerning citizenship puts individuals, especially children, at risk of becoming stateless, [6]. This is due to the fact that there are no adequate regulations that guide the vetting process that certain ethnic groups in Kenya are subjected to. This includes registration offices retaining discretion to request from individuals' documentary proof before issuing documents, including birth certificate and various additional documentations which require repeated trips to various government buildings causing additional travel costs and a prolonged intimidating process.

In Kenya, the known groups of stateless are the Galjeel, Pare and Pemba, [7]. This was the case for stateless persons and persons of undetermined nationality during the 2009 population and housing census of Kenya. This census did not specifically categorize resident persons of unknown nationality in Kenya at that time, hence the stateless population was not clear. However, some studies by the United Nation High Commission for Refugees estimates the stateless population in Kenya as ranging from 18,500 to 20,000 in Kenya, [3].

Despite the attempts to improve the coverage for stateless persons in the 2019 Census, getting the specific groups remained a mirage because the codes or options did not provide for the finer details. Further, it established a much smaller population of 6272. Many of these groups would hide their identities for fear of imagined victimization. Broadly speaking, the Global Action Plan includes actions to resolve existing situations of statelessness; present new cases of statelessness from emerging and better identify and protect stateless persons. The Global Plan to End Statelessness in 10 years requires all states to improve quantitative and qualitative data on stateless populations. The goal specifically requires that quantitative data on stateless populations are publicly available for 150 States and that qualitative analysis of this group is publicly available for at least 120 States, [6].

This study focuses on the stateless persons in Kenya and narrows them down to the Pemba community who is estimated to have a population of about 4000 in Kenya, [8]. It therefore looks into how the Pemba community can be integrated into some of the local communities. Thus, it is important to fully understand the characteristics of the Pemba community and find out if there are any similarities against the surrounding communities using attributes generated from the 2009 Kenya Population and Housing Census with the aim of seeing which local community fits best if they are to be absorbed. To achieve this, a nonparametric ker-

nel discrimination function is developed and used for the classification of the Pemba community into the neighboring local communities. The Characteristics/auxiliary information considered here includes education level and employment status. To determine whether the Pemba community is correctly classified in a particular community, miss-classification rates are computed and compared with other existing classification models.

## 2. Review of Classification through Discriminant Analysis

Application of discriminant analysis has gained interest in various fields of social science, economics, education, finance and engineering. For instance, In routine banking or commercial finance, an officer or analyst may wish to classify loan applicants as low or high credit risks on the basis of the elements of certain accounting statements, [2]. According to [9], he viewed the problem of discriminant analysis as that of assigning an unknown observation to a group with a low error rate. The function or functions used for the assignment may be identical to those used in the multivariate analysis of variance. Also [10], defined discriminant analysis and classification as multivariate techniques concerned with separating distinct sets of objects or observations, and with allocating new objects (observations) to previously defined groups. For instance, in the case of personnel selection the acceptance or rejection of an applicant is frequently based on a number of test scores obtained by the applicant. In all this problems it is assumed that there are two populations, say $P_1$ and $P_2$, one representing the population of individuals fit, and the other the population of individuals unfit for the purpose under consideration. The problem is that of classifying an individual into one of the populations $P_1$ and $P_2$ on the basis of his test scores. Usually, some statistical data from past experience are available which can be utilized in making the classification.

There is a lot of literature where researchers have discussed classification problems extensively and its applications. For instance, discriminant analysis has been applied in classification of students on the basis of their academic performance, [11]. In their research, they used the cumulative results of PRE-ND students of Accountancy and Business Administration department based on the five courses they offered for 2004/2005 academic session. Based on their scores, 78 students were discriminated from Business Administration to Accountancy, and 37 students from Accountancy to Business Administration. In the field of risk analysis, [12] applied discriminant analysis to identify students who might be "At risk" (AR) and "Not At Risk" (NAR). The first group, are students who are in danger of graduating with a poor class of degree, and the second group are those that will graduate with better class of degree within their first two years of study. His analysis successfully classified or predicted 87.5 percent of the graduating students' class of degree. In the education sector, [13] applied discriminant analysis to compare the performance of students who gained admission into the university system through pre-degree programme and those who passed

through the University Matriculation Examination, (UME). It was observed that there is no difference in the performance of UME and predegree students on the average at 5% level of significance.

The gap in the literature cited here is that the researchers relied upon the parametric discriminant methods in the classification problems. Although these methods are conceptually simple and has been used in many application areas, their reliability on the normality assumption limits their performance and application. Furthermore, they are not capable of capturing nonlinearly clustered structures in the data. There is no or little literature that discusses the application of classification in solving the stateless problem that exists globally.

To minimize the failures of the parametric techniques discussed above, this study develops a Robust Nonparametric Kernel discriminant function that will be a better choice whenever a non-linear classification model is needed. This is because Non-parametric estimators are more robust and are useful especially when there exists auxiliary information on finite population parameters which is often used to increase precision of estimators of the parameters, [14].

## 3. Discrimination and Classification

Consider a set of $v$ populations or groups that correspond to density functions $f_1, f_2, \cdots, f_v$. Also consider assigning all the points $x$ from the sample space to one of these groups or densities. The weighted heights of the density functions is used to obtain the Bayes discriminant rule

$$x \text{ is allocated to group } j_0 \text{ if } j_0 = \arg\max_{j\in 1,\cdots,v} \pi_j f_j(x) \tag{1}$$

where $\pi_j$ is the prior probability of drawing from density $f_j$. Enumerating for all $x$ from the sample space, a partition $P = \{P_1, P_2, \cdots, P_v\}$ of the sample space is produced using

$$x \in P_j \text{ if } x \text{ is allocated to group } j$$

The discriminant rule, Equation (1), contains the unknown density functions and the (possibly) unknown prior probabilities. When data is collected, this abstract rule can be modified into a practical one.

The training data $X_j = \{X_{j1}, X_{j2}, \cdots, X_{jn_j}\}$, is collected which is drawn from $f_j$, for $j = 1, 2, \cdots, v$. (The sample sizes $n_j$ are known and non-random).

A priori there is a class structure in the population since it's known which data points are drawn from which density function. From these training data, a practical discriminant rule and subsequent partition can be developed.

Using this discriminant rule/partition, the test data $Y_1, Y_2, \cdots, Y_m$, drawn from

$$f = \sum_{i\in s} y_i + \sum_{j=i}^{v} \pi_j f_j(x)$$

can be classified.

It's not clear here which populations generated which data points.

The usual approach (and the one used in the above example) is to estimate

these density functions (and prior probabilities if needed) and substitute into the discriminant rule. Parametric approaches that are well-known and widely used are linear and quadratic discriminant techniques. However these suffer from the restrictive assumption of normality. With non-parametric discriminant analysis, this assumption can be relaxed and thus be able to tackle more complex cases. The study will focus on kernel methods for discriminant analysis. The monographs [15] [16] [17] (Chapter 7) contain summaries of kernel discriminant analysis while [18] contains more detailed and lengthy expositions on this subject.

### 3.1. Classification of Stateless Persons through Kernel Discriminant Function

Kernel density estimation, [15] [19] is a popular method for nonparametric density estimation, and it has one well known application in kernel discriminant analysis (KDA), [20]. Consider a $J$ class classification problem, if there exist have a training sample $S=\left\{\left(x_i,c_i\right);x_i\in\mathbb{R}^d,C_i\in\left(1,2,\cdots,J\right),i=1,2,\cdots,n\right\}$ of $n$ observations, the kernel estimate for the density function $f_j\left(j=1,2,\cdots,J\right)$ can be expressed as

$$\hat{f}_{jh}\left(x\right)=\frac{1}{nh^d}\sum_{i:c_i=j}K\left\{\frac{1}{h}\left(x-x_i\right)\right\} \tag{2}$$

where $n_j$ is the number of observations from the $j$th class $\sum n_j=n$ $K$ is a $d$-dimensional density function symmetric around 0, and $h$ is the associated smoothing parameter known as the bandwidth. These kernel density estimates are then used to used to construct the proposed kernel discriminant rule (KDR) the proposed classification rule for the stateless persons given by

$$\text{KDR}:\text{is allocated to group } j_0 \text{ if } j_0=\arg\max_{j\in\{1,\cdots,v\}}\hat{\pi}_j\hat{f}_j\left(x,H_j\right) \tag{3}$$

where $\hat{f}_j\left(x,H_j\right)$ is the kernel density estimate corresponding to the $j$th group and where $\pi_j$ is the prior probability of the $j$th group. If these priors are not known, one usually estimates them using training sample proportions $\hat{\pi}_j=\frac{n_j}{n},\left(j=1,2,\cdots,J\right)$ of different groups. Many choices for the kernel function $K$ are available in the literature, [15] [19]. Since the kernel density estimators for discriminant analysis is being used, selection of appropriate bandwidths becomes crucial. One can attempt to find optimal bandwidths for optimal individual kernel density estimates on one hand, while on the other hand, optimal bandwidths which directly optimise the misclassification rate (MR), as [20] attempt for the two can be found.

### 3.2. Misclassification Rate (MR)

This rate is the proportion of points that are assigned to an incorrect group based on a discriminant rule. Then we have

$$\begin{aligned} 1-\mathrm{MR} &= P\left(Y \text{ is classified correctly}\right) \\ &= E_Y\left[1\left\{Y \text{ is classified correctly}\right\}\right] \\ &= E_X\left[E_Y\left[1\left\{Y \text{ is classified correctly}\right\}\right] \mid X_1, X_2, \cdots, X_v\right] \\ &= 1-\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \end{aligned} \tag{4}$$

where $E_Y$ is expectation with respect to $Y$ or $\sum_{j=1}^{v}\pi_j f_j$, and $E_X$ is expectation with respect to $X_1, X_2, \cdots, X_v$ or $\pi_1 f_1, \pi_2 f_2, \pi_v f_v$.

- True positive (TP): Observation is predicted positive and is actually positive.
- False positive (FP): Observation is predicted positive and is actually negative.
- True negative (TN): Observation is predicted negative and is actually negative.
- False negative (FN): Observation is predicted negative and is actually positive.

[18] recommends the former approach for three reasons. First, accurate estimates of the individual density functions are useful in their own right; second, accurate density estimates can be used in other, more complex discriminant problems which look at measures other than the misclassification rate; and third, direct optimisation with respect to a misclassification rate poses many difficult mathematical obstacles.

Whilst we will not use the misclassification rate to select bandwidths, we will still use it as our performance measure of a discriminant rule. So we need to estimate it. The most appropriate estimate depends on whether we have test data or not. If we do, as is the usual case for simulated data, then a simple estimate is obtained by counting the number of $Y_j$ that is assigned to an incorrect group, divided by the total number of data points *m*. On the other hand, if we do not have test data, as is the usual case for real data, then we use the cross validation estimate of MR, as recommended by [15] [18]. This involves leaving out each $X_{ji}$, constructing a corresponding leave-one-out density estimate and subsequent discriminant rule. We then compare the label assigned to $X_{ji}$ based on the leave-one-out discriminant rule to its correct group label. These counts are then summed and divided by *n*.

### 3.3. Algorithm for Kernel Discriminant Classification Rule

The algorithm for the proposed kernel discriminant analysis is given below. The algorithms for linear and quadratic discriminant analysis are similar except that any kernel methods are replaced by the appropriate parametric methods. We put these algorithms into practice with both simulated and real data.

1) For each training sample $X_j = \left\{X_{j1}, X_{j2}, \cdots, X_{jn_j}\right\}$, $j = 1, 2, \cdots, v$, compute a kernel density estimate

$$\hat{f}\left(x; H_j\right) = n_j^{-1}\sum_{i=1}^{n_j} K_{H_j}\left(x - X_{ji}\right) \tag{5}$$

We can use any sensible bandwidth selector $H_j$

2) If prior probabilities are available, then use them. Otherwise estimate them using the training sample proportions $\hat{\pi}_j = n_j / n$.

3)

a) Allocate test data points $Y_1, Y_2, \cdots, Y_m$ according to KDR/Equation (3) or,

b) Allocate all points $x$ from the sample space according to KDR/Equation (3).

4)

a) If we have test data then the estimate of the misclassification rate is

$$\hat{\mathrm{MR}} = 1 - m^{-1} \sum_{k=1}^{v} 1\{Y_K \text{ is classified correctly using KDR}\} \tag{6}$$

b) If we do not have test data the cross validation estimate of the misclassification rate is

$$\hat{\mathrm{MR}}_{\mathrm{CV}} = 1 - n^{-1} \sum_{j=1}^{v} \sum_{i=1}^{n_j} 1\{X_{ji} \text{ is classified correctly using KDR}_{ji}\} \tag{7}$$

where $\mathrm{KDR}_{ji}$ is similar to KDR except that $\hat{f}_j(.; H_j)$ and $\hat{\pi}_j$ are replaced by their leave one out estimates obtained by removing $X_{ji}$ that is $\hat{\pi}_{ji} = (n_j - 1)/n$ and

$$\hat{f}_{j,-i}(x; H_j) = (n_j - 1)^{-1} \sum_{i'=1, i' \neq 1}^{n_j} K_{H_{j,-i}}(x - X_{ji'}) \tag{8}$$

That is, we repeat step 3 to classify all $X_{ji}$ using $\mathrm{KDR}_{ji}$.

## 4. Emperical Study

For the real data, we are using data from Kenya National Bureau of Statistics obtained from the 2009 Census. The data consist of tribes living in the coastal region of Kenya especially the Kilifi county and various characteristics associated to them such as Education level Religion, Building material, waste disposal, source of water and employment status. The study aims to classify these communities using the characteristics observed amongst them and obtain the misclassification error which is the error that the community is classified in the wrong group. In addition, the study aims at using this information to classify the Pemba community which has been stateless for long time and use this information to advice the policy makers to consider integrating the Pemba people into the identified community/s. This will help to inform on the classification decision on any emerging tribe in the coastal region whose is not known but possess similar characteristics. Due to the challenges of insufficient data in the database on the Pemba Community, the only data available for use is based on the characteristics such as level of education and employment. We apply non parametric discriminant analysers and compare their performance with the parametric methods.

As shown in **Figure 1**, there are about 10 communities in Kilifi County with a population of about 1.02 million people which are neighboring the Pemba community which has an estimated population of over 2000 people and has been stateless for a long time since Kenya got Independent. Although some Pemba

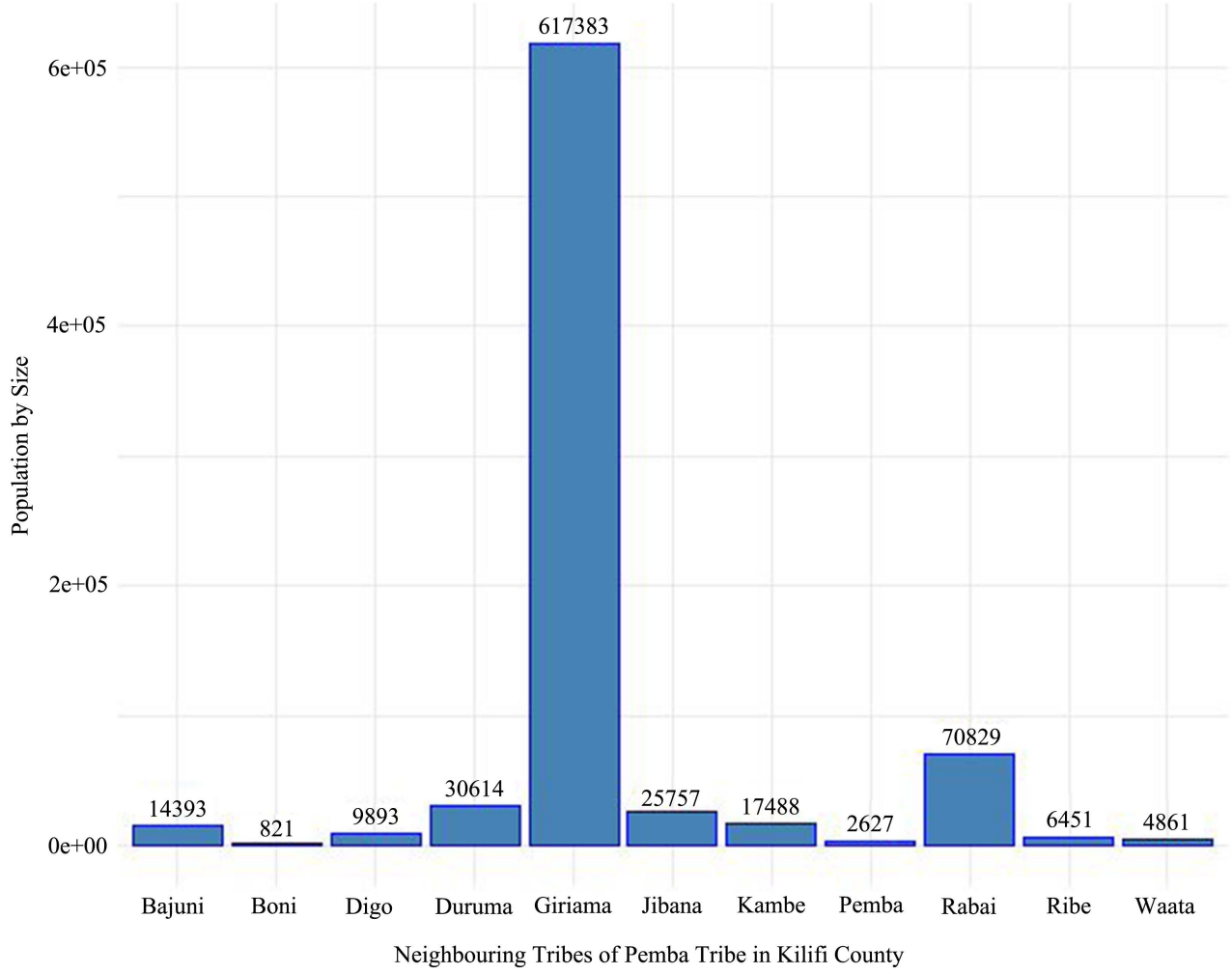


**Figure 1.** Distributions of the tribes neighboring the pemba community in Kilifi County.

were issued with IDs in Kenya, most of the IDs were withdrawn or not renewed with the change in administration and legislation. After their identity documents were withdrawn in the 1980s and late 1990s, many Pemba were asked to leave the country but they would spend days hiding in the bushes until the situation seem calm enough for them to return. This community, who are mainly fishermen by trade, cannot obtain a fishing license and have no access to relief food during emergencies and they cannot take even enjoy of banking services.

To analyse this data and perform a classification, a sample of 3000 observation was taken using Stratified simple random sampling technique where the tribes were treated as the stratas. The proportional allocation technique was used to obtain a sample from each tribe to ensure equal representation in the each tribe in the study. The sample data was then divided in two parts which 66% being used to train various classifiers used to perform the classification and 34% used for the testing and classification of the various communities into specific tribes.

A comparison is conducted by examining the performance of the following

discriminant analysers:

1) Linear discriminant (LD).
2) Quadratic discriminant (QD).
3) Kernel discrinant with 2-stage AMSE diagonal bandwidth matrices (KDD2).
4) Kernel discrinant with 2-stage SAMSE full bandwidth matrices (KDS2).
5) Kernel discrinant with 1-stage SCV full bandwidth matrices (KDSC).

The R code for kernel discriminant analysers is based on the bandwidth matrix selection and density functions in the *ks* library. The R code for LDA and QDA are supplied within the MASS library in the R software by the function *lda*() and *qda*() respectively.

### 4.1. Misclassification Rates for Stateless Communities in Kenya

In the first analysis, we use the training data to train the model and use the same training data as a test data to see how the model performs.

The misclassification rate rates within the groups are given in Table 1 and Table 2. From these results it can be observed that, Kernel discriminant anaysers are more efficient than parametric ones.

In the second analysis, we use the training data to train the model and an independent data as a test data to see how the model performs. From the results in Table 3, the cross validation misclassification rates for the kernel discriminants are KDD2: 0.5375, KDS2: 0.4875 and KDSC: 0.5689. For the parametric discriminants, they are LD: 0.7625 and QD: 0.7000. It can be observed that the kernel methods, with appropriately chosen bandwidth matrices, outperform the parametric methods; and that the kernel methods with full bandwidth matrices outperform those with diagonal bandwidth matrices.

In some instances accuracy or misclassification error can be misleading if used with imbalanced datasets, and therefore there are other performance metrics based on confusion matrix which can be useful for evaluating performance. These performance measures include Sensitivity, Specificity, Precision, Recalls and F1. Precision or the positive predictive value, is the fraction of positive values out of the total predicted positive instances. In other words, precision is the proportion of positive values that were correctly identified; Sensitivity, recall, or the TP rate (TPR) is the fraction of positive values out of the total actual positive

**Table 1.** Misclassification rates for various discrimant analyser using training data as a test data.

| Method | Misclassification Rate |
|---|---|
| KDDS2 | 0.0813 |
| KDS2 | 0.0750 |
| KDSC | 0.0875 |
| LD | 0.3625 |
| QD | 0.1563 |

**Table 2.** Misclassification rates for each group for various discrimant analyser using training data as a test data.

| Tribe | Misclassification rate | | | | |
|---|---|---|---|---|---|
| | KDD2 | KDS2 | KDSC | LD | QD |
| Bajuni | 0.0000 | 0.0000 | 0.0000 | 0.9438 | 0.9375 |
| Boni | 0.3000 | 0.3000 | 0.3000 | 0.9625 | 0.9688 |
| Digo | 0.0.000 | 0.0000 | 0.0000 | 0.9438 | 0.9438 |
| Duruma | 0.0000 | 0.0000 | 0.0000 | 0.9375 | 0.9375 |
| Giriama | 0.9000 | 0.80000 | 0.9000 | 0.9938 | 0.9938 |
| Jibana | 0.0000 | 0.0000 | 0.0000 | 0.9500 | 0.9375 |
| Kambe | 0.0000 | 0.0000 | 0.0000 | 0.9438 | 0.9438 |
| Pemba | 0.1000 | 0.100 | 0.2000 | 0.9875 | 0.9688 |
| Rabai | 0.0000 | 0.0000 | 0.0000 | 0.9875 | 0.9500 |
| Ribe | 0.0000 | 0.0000 | 0.0000 | 0.9500 | 0.9375 |
| Wataa | 0.0000 | 0.00000 | 0.0000 | 0.9500 | 0.9375 |

**Table 3.** Misclassification rates for various discrimant analyser using independent test data.

| Method | Misclassification Rate | Kappa |
|---|---|---|
| KDD2 | 0.5375 | 0.2806 |
| KDS2 | 0.4875 | 0.2921 |
| KDSC | 0.56875 | 0.2743 |
| LD | 0.7625 | 0.2544 |
| QD | 0.7000 | 0.2484 |

instances (*i.e.* the proportion of actual positive cases that are correctly identified, while Specificity gives the fraction of negative values out of the total actual negative instances. In other words, it is the proportion of actual negative cases that are correctly identified. The FP rate is given by (1 − specificity). The F1 score, F score, or F measure is the harmonic mean of precision and sensitivity it gives importance to both factors. **Table 4** gives these performance measures for each tribe based on different classifiers. It can be observed that the Kernel discriminant classifiers outperform parametric classifiers when the appropriate bandwidth matrix is chosen as they show high values of precision, sensitivity, specificity and F1 across the tribes.

### 4.2. Classification of the Stateless Pemba Community

The main objective of this study was find which neighboring local community in Kilifi County that the Pemba community which have lived for long in a stateless nature can be integrated into so that they can be recognized as Kenyans and be issued with the National Identification Number so that they can be able to access

**Table 4.** Classification performance of four classification models based on the data on the stateless communities.

| | | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KDD2 | Sensitivity | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.54904 | 1.00000 | 0.00000 | 0.6259 | 0.29293 | 0.00000 | 0.00000 |
| | Specificity | 0.95419 | 0.96686 | 0.96491 | 0.93275 | 0.97095 | 0.93659 | 0.95322 | 0.94363 | 0.94498 | 0.96686 | 0.96387 |
| | Precision | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.98402 | 0.01515 | 0.00000 | 0.63504 | 0.3625 | 0.00000 | 0.00000 |
| | Recall | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.54904 | 1.00000 | 0.00000 | 0.6259 | 0.29293 | 0.00000 | 0.00000 |
| | F1 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.70482 | 0.02985 | 0.00000 | 0.63043 | 0.32402 | 0.00000 | 0.00000 |
| KDS2 | Sensitivity | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.54962 | 1.00000 | 0.00000 | 0.61702 | 0.29167 | 0.00000 | 0.00000 |
| | Specificity | 0.95419 | 0.96686 | 0.96491 | 0.93275 | 0.975 | 0.93659 | 0.95322 | 0.9435 | 0.94409 | 0.96686 | 0.96387 |
| | Precision | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.9863 | 0.01515 | 0.00000 | 0.63504 | 0.35 | 0.00000 | 0.00000 |
| | Recall | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.54962 | 1.00000 | 0.00000 | 0.61702 | 0.29167 | 0.00000 | 0.00000 |
| | F1 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.70588 | 0.02985 | 0.00000 | 0.6259 | 0.31818 | 0.00000 | 0.00000 |
| KDSC | Sensitivity | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.54753 | 0.00000 | 0.00000 | 0.61111 | 0.27174 | 0.00000 | 0.00000 |
| | Specificity | 0.95419 | 0.96686 | 0.96491 | 0.93275 | 0.97468 | 0.93567 | 0.95322 | 0.94444 | 0.94111 | 0.96686 | 0.9639 |
| | Precision | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.9863 | 0.00000 | 0.00000 | 0.64234 | 0.31250 | 0.00000 | 0.00000 |
| | Recall | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.54753 | 0.00000 | 0.00000 | 0.61111 | 0.27174 | 0.00000 | 0.00000 |
| | F1 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.70416 | 0.00000 | 0.00000 | 0.62633 | 0.2907 | 0.00000 | 0.00000 |
| QDA | Sensitivity | 0.14286 | 0.00000 | 0.00000 | 0.10000 | 0.56601 | 0.14894 | 0.21429 | 0.65289 | 0.25275 | 0.07692 | 0.07692 |
| | Specificity | 0.95553 | 0.96654 | 0.96457 | 0.9334 | 0.82289 | 0.93973 | 0.95792 | 0.93591 | 0.93904 | 0.96742 | 0.96446 |
| | Precision | 0.04255 | 0.00000 | 0.00000 | 0.02899 | 0.8516 | 0.10606 | 0.12500 | 0.57664 | 0.2875 | 0.02941 | 0.02703 |
| | Recall | 0.14286 | 0.00000 | 0.00000 | 0.10000 | 0.56601 | 0.14894 | 0.21429 | 0.65289 | 0.25275 | 0.07692 | 0.07692 |
| | F1 | 0.06557 | 0.00000 | 0.00000 | 0.04494 | 0.68004 | 0.12389 | 0.15789 | 0.6124 | 0.26901 | 0.04255 | 0.04000 |
| LDA | Sensitivity | 0.07143 | 0.00000 | 0.00000 | 0.09524 | 0.56476 | 0.15556 | 0.15152 | 0.64407 | 0.25275 | 0.11111 | 0.06667 |
| | Specificity | 0.95455 | 0.96670 | 0.96453 | 0.93333 | 0.82597 | 0.93986 | 0.95670 | 0.93282 | 0.93904 | 0.96755 | 0.96439 |
| | Precision | 0.02128 | 0.00000 | 0.00000 | 0.02899 | 0.85616 | 0.10606 | 0.10417 | 0.55474 | 0.2875 | 0.02941 | 0.02703 |
| | Recall | 0.07143 | 0.00000 | 0.00000 | 0.09524 | 0.56476 | 0.15556 | 0.15152 | 0.64407 | 0.25275 | 0.11111 | 0.06667 |
| | F1 | 0.03279 | 0.00000 | 0.00000 | 0.04444 | 0.68058 | 0.12613 | 0.12346 | 0.59608 | 0.26901 | 0.04651 | 0.03846 |

Services from the Government just like any other Kenyans without discrimination. Such activities include access to some basic rights and services such as acquisition of birth certificates, education, formal employment, financial services, for example, opening a bank account, in some cases health care, health insurance services, and to play in sports at national and international levels. The neighboring local communities the study is seeking to integrate Pemba Community into includes the Bajuni, Boni, Digo, Duruma, Giriama, Jibana, Kambe, Rabai, Ribe and Waata community living in Kilifi county where majority of the Pemba community are found.

The results in Table 5, Table 7, Table 9 and Table 11 present the confusion matrix for the classification of the communities using the Kernel discriminant classifier KDD2, KDSC and the Quadratic and Linear discriminant classifier re-

spectively. From these results, the KDD2 classifier apart from truly classifying Pemba community as Pemba, it also classified them into other tribes with 29 people being classified as Giriama, 87 as Pemba and 21 people as Rabai. The KDSC classifier classified 29 people as Giriama, 88 as Pemba and 20 as Rabai. The QDA classifier classified majority 20 as Giriama, 79 as Pemba and 20 as Rabai and the LDA classifier classified the Pemba people with 22 being classified as Giriama, 76 as true Pemba and 19 as the Rabai. From this finding it can be observed that, based on certain similarities that exists in this communities, the Pemba community can be classified as Giriama because they seem to have a strong link with them. The next community that they can be classified as is the Rabai community (Tables 5-12).

**Table 5.** The confusion matrix of the communities in Kilifi County classified based on KDD2 classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 0 | 0 | 0 | 0 | 45 | 0 | 0 | 0 | 2 | 0 | 0 | 47 |
| Boni | 0 | 0 | 0 | 0 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 34 |
| Digo | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 36 |
| Duruma | 0 | 0 | 0 | 0 | 67 | 0 | 0 | 0 | 1 | 0 | 1 | 69 |
| Giriama | 0 | 0 | 0 | 0 | 431 | 0 | 0 | 1 | 6 | 0 | 0 | 438 |
| Jibana | 0 | 0 | 0 | 0 | 26 | 1 | 0 | 17 | 22 | 0 | 0 | 66 |
| Kambe | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 10 | 18 | 0 | 0 | 48 |
| Pemba | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 87 | 21 | 0 | 0 | 137 |
| Rabai | 0 | 0 | 0 | 0 | 28 | 0 | 0 | 23 | 29 | 0 | 0 | 80 |
| Ribe | 0 | 0 | 0 | 0 | 33 | 0 | 0 | 0 | 0 | 0 | 1 | 34 |
| Waata | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 1 | 0 | 0 | 0 | 37 |

**Table 6.** The Confusion matrix of the proportion of the communities being classified correctly into a particular community based on KDD2 classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.95745 | 0.00000 | 0.00000 | 0.00000 | 0.04255 | 0.00000 | 0.00000 |
| Boni | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 1.0000 | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.00000 | 0.00000 |
| Digo | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 1.0000 | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.00000 | 0.00000 |
| Duruma | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.97101 | 0.00000 | 0.00000 | 0.00000 | 0.01449 | 0.00000 | 0.01449 |
| Giriama | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.98402 | 0.00000 | 0.00000 | 0.00228 | 0.01370 | 0.00000 | 0.00000 |
| Jibana | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.39394 | 0.01515 | 0.00000 | 0.25758 | 0.33333 | 0.00000 | 0.00000 |
| Kambe | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.41667 | 0.00000 | 0.00000 | 0.20833 | 0.37500 | 0.00000 | 0.00000 |
| Pemba | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.21168 | 0.00000 | 0.0000 | 0.63504 | 0.15328 | 0.00000 | 0.0000 |
| Rabai | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.35000 | 0.00000 | 0.0000 | 0.28750 | 0.36250 | 0.00000 | 0.0000 |
| Ribe | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.97059 | 0.00000 | 0.0000 | 0.00000 | 0.0000 | 0.00000 | 0.02941 |
| Waata | 0.00000 | 0.00000 | 0.00000 | 0.0000 | 0.97297 | 0.00000 | 0.0000 | 0.02703 | 0.00000 | 0.0000 | 0.00000 |

**Table 7.** The confusion matrix of the communities in Kilifi County classified based on KDSC classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 0 | 0 | 0 | 0 | 46 | 0 | 0 | 0 | 1 | 0 | 0 | 47 |
| Boni | 0 | 0 | 0 | 0 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 34 |
| Digo | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 36 |
| Duruma | 0 | 0 | 0 | 0 | 67 | 0 | 0 | 0 | 1 | 0 | 1 | 69 |
| Giriama | 0 | 0 | 0 | 0 | 432 | 0 | 0 | 1 | 5 | 0 | 0 | 438 |
| Jibana | 0 | 0 | 0 | 0 | 26 | 0 | 0 | 18 | 22 | 0 | 0 | 66 |
| Kambe | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 10 | 18 | 0 | 0 | 48 |
| Pemba | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 88 | 20 | 0 | 0 | 137 |
| Rabai | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 25 | 25 | 0 | 0 | 80 |
| Ribe | 0 | 0 | 0 | 0 | 33 | 0 | 0 | 1 | 0 | 0 | 0 | 34 |
| Waata | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 1 | 0 | 0 | 0 | 37 |

**Table 8.** The Confusion matrix of the proportion of the communities being classified correctly into a particular community based on KDSC classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.97872 | 0.00000 | 0.00000 | 0.00000 | 0.02128 | 0.00000 | 0.00000 |
| Boni | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 1.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| Digo | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 1.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| Duruma | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.97101 | 0.00000 | 0.00000 | 0.00000 | 0.01449 | 0.00000 | 0.01449 |
| Giriama | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.9863 | 0.00000 | 0.00000 | 0.00228 | 0.01142 | 0.00000 | 0.00000 |
| Jibana | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.39394 | 0.00000 | 0.00000 | 0.27273 | 0.33333 | 0.00000 | 0.00000 |
| Kambe | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.41667 | 0.00000 | 0.00000 | 0.20833 | 0.37500 | 0.00000 | 0.00000 |
| Pemba | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.21168 | 0.00000 | 0.00000 | 0.64234 | 0.14599 | 0.00000 | 0.00000 |
| Rabai | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.37500 | 0.00000 | 0.00000 | 0.31250 | 0.31250 | 0.00000 | 0.00000 |
| Ribe | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.97059 | 0.00000 | 0.00000 | 0.02941 | 0.00000 | 0.00000 | 0.00000 |
| Waata | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.97297 | 0.00000 | 0.00000 | 0.02703 | 0.00000 | 0.00000 | 0.00000 |

**Table 9.** The confusion matrix of the communities in Kilifi County classified based on QDA classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 2 | 0 | 1 | 1 | 33 | 1 | 0 | 3 | 5 | 1 | 0 | 47 |
| Boni | 0 | 0 | 0 | 0 | 31 | 0 | 1 | 0 | 1 | 1 | 0 | 34 |
| Digo | 1 | 0 | 0 | 1 | 31 | 1 | 0 | 0 | 1 | 0 | 1 | 36 |
| Duruma | 0 | 0 | 1 | 2 | 58 | 1 | 0 | 2 | 2 | 0 | 3 | 69 |
| Giriama | 8 | 7 | 7 | 9 | 373 | 5 | 3 | 5 | 7 | 10 | 4 | 438 |

**Continued**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jibana | 1 | 0 | 0 | 0 | 18 | 7 | 9 | 13 | 18 | 0 | 0 | 66 |
| Kambe | 0 | 1 | 0 | 0 | 13 | 8 | 6 | 6 | 13 | 0 | 1 | 48 |
| Pemba | 0 | 0 | 0 | 3 | 20 | 10 | 4 | 79 | 20 | 0 | 1 | 137 |
| Rabai | 1 | 1 | 1 | 2 | 23 | 12 | 5 | 11 | 23 | 0 | 1 | 80 |
| Ribe | 1 | 1 | 0 | 1 | 26 | 2 | 0 | 1 | 0 | 1 | 1 | 34 |
| Waata | 0 | 0 | 0 | 1 | 33 | 0 | 0 | 1 | 1 | 0 | 1 | 37 |

**Table 10.** The confusion matrix of the proportion of the communities being classified correctly into a particular community based on QDA classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 0.04260 | 0.00000 | 0.02130 | 0.02130 | 0.70210 | 0.02130 | 0.00000 | 0.06380 | 0.10640 | 0.02130 | 0.00000 |
| Boni | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.91180 | 0.00000 | 0.02940 | 0.00000 | 0.02940 | 0.0294 | 0.00000 |
| Digo | 0.02780 | 0.00000 | 0.00000 | 0.02780 | 0.86110 | 0.02780 | 0.00000 | 0.00000 | 0.0278 | 0.00000 | 0.0278 |
| Duruma | 0.00000 | 0.00000 | 0.01450 | 0.02900 | 0.84060 | 0.0145 | 0.00000 | 0.02900 | 0.02900 | 0.00000 | 0.0435 |
| Giriama | 0.01830 | 0.01600 | 0.01600 | 0.02060 | 0.85160 | 0.01140 | 0.00690 | 0.01140 | 0.01600 | 0.02280 | 0.00910 |
| Jibana | 0.01520 | 0.00000 | 0.00000 | 0.00000 | 0.27270 | 0.10610 | 0.13640 | 0.19700 | 0.2727 | 0.00000 | 0.00000 |
| Kambe | 0.00000 | 0.02080 | 0.00000 | 0.00000 | 0.27080 | 0.16670 | 0.12500 | 0.12500 | 0.27080 | 0.00000 | 0.02080 |
| Pemba | 0.00000 | 0.00000 | 0.00000 | 0.02190 | 0.14600 | 0.07300 | 0.02920 | 0.57660 | 0.14600 | 0.00000 | 0.00730 |
| Rabai | 0.01250 | 0.01250 | 0.01250 | 0.02500 | 0.28750 | 0.15000 | 0.06250 | 0.13750 | 0.2875 | 0.00000 | 0.0125 |
| Ribe | 0.02940 | 0.0294 | 0.00000 | 0.02940 | 0.76470 | 0.0588 | 0.00000 | 0.0294 | 0.00000 | 0.02940 | 0.02940 |
| Waata | 0.00000 | 0.00000 | 0.00000 | 0.02700 | 0.89190 | 0.00000 | 0.00000 | 0.02700 | 0.02700 | 0.00000 | 0.02700 |

**Table 11.** The confusion matrix of the communities in Kilifi County classified based on LDA classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 1 | 0 | 1 | 1 | 34 | 1 | 0 | 3 | 5 | 1 | 0 | 47 |
| Boni | 0 | 0 | 0 | 0 | 31 | 0 | 1 | 0 | 1 | 0 | 1 | 34 |
| Digo | 1 | 0 | 0 | 1 | 32 | 1 | 0 | 0 | 0 | 0 | 1 | 36 |
| Duruma | 0 | 0 | 2 | 2 | 57 | 1 | 0 | 2 | 2 | 0 | 3 | 69 |
| Giriama | 8 | 3 | 7 | 10 | 375 | 3 | 5 | 6 | 9 | 7 | 5 | 438 |
| Jibana | 1 | 0 | 0 | 0 | 19 | 7 | 10 | 13 | 16 | 0 | 0 | 66 |
| Kambe | 0 | 0 | 0 | 1 | 13 | 8 | 5 | 5 | 15 | 0 | 1 | 48 |
| Pemba | 0 | 0 | 0 | 2 | 22 | 10 | 7 | 76 | 19 | 0 | 1 | 137 |
| Rabai | 2 | 1 | 0 | 2 | 23 | 12 | 5 | 11 | 23 | 0 | 1 | 80 |
| Ribe | 1 | 1 | 0 | 1 | 26 | 2 | 0 | 1 | 0 | 1 | 1 | 34 |
| Waata | 0 | 0 | 1 | 1 | 32 | 0 | 0 | 1 | 1 | 0 | 1 | 37 |

**Table 12.** The confusion matrix of the proportion of the communities being classified correctly into a particular community based on LDA classifier.

| | Bajuni | Boni | Digo | Duruma | Giriama | Jibana | Kambe | Pemba | Rabai | Ribe | Waata |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajuni | 0.0213 | 0.0000 | 0.0213 | 0.0213 | 0.7234 | 0.0213 | 0.0000 | 0.0638 | 0.1064 | 0.0213 | 0.0000 |
| Boni | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.9118 | 0.0000 | 0.0294 | 0.0000 | 0.0294 | 0.0000 | 0.0294 |
| Digo | 0.0278 | 0.0000 | 0.0000 | 0.0278 | 0.8889 | 0.0278 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0278 |
| Duruma | 0.0000 | 0.0000 | 0.0290 | 0.0290 | 0.8261 | 0.0145 | 0.0000 | 0.0290 | 0.0290 | 0.0000 | 0.0435 |
| Giriama | 0.0183 | 0.0069 | 0.0160 | 0.0228 | 0.8562 | 0.0069 | 0.0114 | 0.0137 | 0.0206 | 0.0160 | 0.0114 |
| Jibana | 0.0152 | 0.0000 | 0.0000 | 0.0000 | 0.2879 | 0.1061 | 0.1515 | 0.1970 | 0.2424 | 0.0000 | 0.0000 |
| Kambe | 0.0000 | 0.0000 | 0.0000 | 0.0208 | 0.2708 | 0.1667 | 0.1042 | 0.1042 | 0.3125 | 0.0000 | 0.0208 |
| Pemba | 0.0000 | 0.0000 | 0.0000 | 0.0146 | 0.1606 | 0.0730 | 0.0511 | 0.5547 | 0.1387 | 0.0000 | 0.0073 |
| Rabai | 0.0250 | 0.0125 | 0.0000 | 0.0250 | 0.2875 | 0.1500 | 0.0625 | 0.1375 | 0.2875 | 0.0000 | 0.0125 |
| Ribe | 0.0294 | 0.0294 | 0.0000 | 0.0294 | 0.7647 | 0.0588 | 0.0000 | 0.0294 | 0.0000 | 0.0294 | 0.0294 |
| Waata | 0.0000 | 0.0000 | 0.0270 | 0.0270 | 0.8649 | 0.0000 | 0.0000 | 0.0270 | 0.0270 | 0.0000 | 0.0270 |

## 5. Conclusions and Recommendations

### 5.1. Conclusions

The main objective of this paper was to develop a nonparametric discriminant classifier and use it to find which neighboring local community in Kilifi County that the Pemba community which has lived for long in a stateless nature can be integrated into so that they can be recognized as Kenyans and hence live like any other Kenyan Citizen and enjoy.

From the results, the following observations and conclusions have been made:

1) Classification of stateless communities in Kenya can be done using the Kernel discrimination classification methods to find which local communities they can be integrated into.

2) The Nonparametric Kernel Discriminant Classifiers; KDD2 classifier apart from truly classifying Pemba community as Pemba also classified them into other tribes with 29 people being classified as Giriama, 87 as Pemba and 21 people as Rabai. The KDSC classifier classified 29 people as Giriama, 88 as Pemba and 20 as Rabai. The Parametric discriminant classifiers; QDA classifier classified majority of the Pemba people, 20 as Giriama, 79 as Pemba and 20 as Rabai while the LDA classifier classified the Pemba people with 22 being classified as Giriama, 76 as true Pemba and 19 as the Rabai.

3) Based on certain similarities in characteristics that exist in the communities that surround the Pemba Community, the Pemba community can be classified as Giriama in which they seem to have a strong link. The alternative local community that could have Pemba integrated is the Rabai Community.

### 5.2. Recommendations

The study recommends the use of Kernel discriminant technique to classify the stateless communities in Kenya e.g. Pemba. This approach can be extended to similar groups across the world. This will go a long way in achieving UNHCR recommendation of finding a solution on how to recognize the stateless communities and register them as citizens. In addition to this, the study also recommends more data on various dimensions to be collected on the stateless peoples which seem to have been excluded in the census of 2009 conducted by the Kenya so as to allow for more analyses and improve the efficiency of the results obtained. Lastly, the study recommends other classification techniques which can handle the high dimensional spaces such Neural Networks to be considered in the future studies so as to see if efficiency of classification can be improved.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Milbrandt, J. (2011) Stateless. *Cardozo International & Comparative Law Review*, **20**, 75-103.

[2] Manly, M. and Van Waas, L. (2014) The State of Statelessness Research: A Human Rights Imperative. *Tilburg Law Review*, **19**, 3-10. https://doi.org/10.1163/22112596-01902029

[3] Sutton, N.L.J. (2018) Statelessness and the Rights of Children in Kenya and South Africa: A Human Rights Perspective. Master's Thesis, University of the Western Cape, Cape Town.

[4] Muimi, A.M. (2021) Generational Statelessness and Rights a Case of Children from the Nubian Community in Kenya. Strathmore University, Nairobi.

[5] Abuya, E.O. (2021) Registering Persons at Risk of Statelessness in Kenya: Solutions or Further Problems? In: Bloom, T. and Kingston, L.N., Eds., *Statelessness, Governance, and the Problem of Citizenship*, Manchester University Press, Manchester, 251-263.

[6] United Nations High Commissioner for Refugees (2014) Global Action Plan to End Statelessness 2014-2024. United Nations High Commissioner for Refugees, Geneva.

[7] Bosire, C. (2017) Statelessness in a State? *SSRN*, Article ID: 3079430. https://doi.org/10.2139/ssrn.3079430

[8] Kenya National Bureau of Statistics (2009) Kenya Population and Housing Census. Kenya National Bureau of Statistics, Nairobi.

[9] Lachenbruch, P.A. (1974) Discriminant Analysis When the Initial Samples Are Misclassified II: Non-Random Misclassification Models. *Technometrics*, **16**, 419-424. https://doi.org/10.1080/00401706.1974.10489211

[10] Arnold Johnson, R., Wichern, D.W., *et al.* (2014) Applied Multivariate Statistical Analysis. Vol. 6, Pearson, London.

[11] Usoro, A.E. (2006) Discriminant Analysis and Its Application to the Classification of Students on the Basis of Their Academic Performances. *Research Journal of Physical Sciences*, **2**, 53-55.

[12] Erimafa, J.T., Iduseri, A. and Edokpa, I.W. (2009) Application of Discriminant Analysis to Predict the Class of Degree for Graduating Students in a University System. *International Journal of Physical Sciences*, **4**, 16-21.

[13] Thomas, A.U. and Pascal, B.N. (2013) A Comparison of Theperformance of Students in Pre-Degree and University Matriculation Examination Classes in a University System: Adiscriminant Analysis Approach. *International Journal of Advancement in Research and Technology*, **2**, 60-70.

[14] Cochran, W.G. (2007) Sampling Techniques. John Wiley & Sons, Hoboken.

[15] Silverman, B.W. (1998) Density Estimation for Statistics and Data Analysis. Routledge, New York. https://doi.org/10.1201/9781315140919

[16] Scott, D.W. (1991) Feasibility of Multivariate Density Estimates. *Biometrika*, **78**, 197-205. https://doi.org/10.1093/biomet/78.1.197

[17] Simonoff, J.S. (1996) Further Applications of Smoothing. In: *Smoothing Methods in Statistics*, Springer, New York, 252-274. https://doi.org/10.1007/978-1-4612-4026-6_7

[18] Hand, D.J. (1982) Kernel Discriminant Analysis. John Wiley & Sons, Inc., Somerset, 264.

[19] Scott, D.W. (1992) Multivariate Density Estimation: Theory, Practice and Visualisation. John Willey and Sons, Inc., New York.

[20] Hall, P. and Wand, M.P. (1988) On Nonparametric Discrimination Using Density Differences. *Biometrika*, **75**, 541-547. https://doi.org/10.1093/biomet/75.3.541