CLASSIFICATION OF STATELESS COMMUNITIES USING A ROBUST NONPARAMETRIC KERNEL DISCRIMINANT FUNCTION

MACDONALD GEORGE OBUDHO

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2023

Classification of Stateless Communities Using a Robust Nonparametric Kernel Discriminant Function

Macdonald George Obudho

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Applied Statistics of the Jomo Kenyatta University of Agriculture and Technology

2023

DECLARATION

This Thesis is my original work and has not been presented for a degree in ay other University.

Signature.. Date...............................

Macdonald George Obudho

This Thesis has been submitted for examination with our approval as University Supervisors.

Signature.. Date...............................

Prof. Romanus Odhiambo Otieno

JKUAT, Kenya

Signature.. Date...............................

Prof. George Otieno Orwa

JKUAT, Kenya

DEDICATION

To my beloved wife, Carolyne Achieng and my caring mother; Elizabeth Nyar Osula. I will not forget my grandfather; the late Great Ochola Owuondo, my uncle and mentor; the late Prof. Laban Ogallo. My paternal grandmother Priskila, who served as my mother during my childhood is honoured as a major motivator to do this work. I hope that this work will inspire my dear children; Candy, Denzel, Sophie and Noel to work harder in their academic world and make greater achievements.

ACKNOWLEDGMENT

Foremost, I thank the Almighty God for the gift of life that He has offered to me for those many years and the chance to go through Basic Education despite numerous challenges during my childhood. I then sincerely thank my two supervisors; Prof. Romanus Odhiambo Otieno and Prof. George Otieno Orwa for their support and guidance in coming up with this Thesis. Last but not least is to thank my PhD colleague Festus Were who offered peer support during the long journey that we walked together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

AMSE	Asymptotic mean Square Error
AMISE	Asymptotic Mean Integrated Squared Error
AR	At Risk
CDF	Cumulative Density Function
DA	Discriminant Analysis
DFA	Discriminant Function Analysis
DRC	Democratic Republic of Congo
FDA	Fisher's Discriminant Analysis
GIGO	Garbage In Garbage Out
ID	Identification
KCC	Kernel Canonical Correclations
KCCA	Kernel Canonical Correclation Analysis
KDA	Kernel Discriminant Analysis
KDR	Kernel Discriminant Rule
KDD2	Kernel Discrinant function with AMSE bandwidth
KDE	Kernel Density Estimator
KDFA	Kernel Discriminant Function Analysis
KDS2	Kernel Discrinant function with SAMSE bandwidth
KDSC	Kernel Discrinant function with Smooth cross validation bandwidth
KLR	Kernel Logistic Regression
KNBS	Kenya National Bureau of Statistics
KNN	K-nearest neighbour
KPCA	Kernel Principal Component Analysis
LD	Linear Discriminant
LDA	Linear Discriminant Analysis
LKNDA	Local Kernel Nonparametric Discriminant Analysis
LLDA	Local Linear Discriminant Analysis

LNOB	Leaving No One Behind
MISE	Mean Intergrated Squared Error
MR	Misclassification Rate
NAR	Not At Risk
NDA	Nonparametric Discriminant Analysis
NFA	Nonparametric Feature Analysis
NNDA	Nearest Neigbour Discriminant Analysis
PCA	Principal Component Analysis
QDA	Quadratic Discriminant Analysis
RBF	Radial Basis Function
RDA	Regularized Discriminant Analysis
RMSE	Root Mean Squared Error
SAMSE	Sum of Asymptotic mean Square Error
SCV	Smoothed Cross Validation
SDGs	Sustainable Development Goals
SSS	Small sample size
SVMs	Support Vector Machines
UDHR	Universal Declaration of Human Rights
UME	University Matriculation Examination
UNHCR	United Nations High Commissioner for Refugees

ABSTRACT

The main objective of this study was to classify the stateless communities using a Robust Nonparametric Kernel Discriminant Function. A Robust Nonparametric Kernel Discriminant Function has therefore been developed by modifying the traditional using Bayes Discriminant Rule with a Nonparametric Kernel Discriminant Function. A suitable Kernel method was carefully chosen and a series of bandwidths were tested to get what could work best for our model. The study also estimated the Classification Rates of the developed function as a measure of its Robustness. The function was compared with parametric functions such as Linear Discriminant Function and Quadratic Discriminant Function through a simulation study. The result has been applied in classifying the stateless communities. As of today, the Pemba people in Kenya are among a number of other communities in the world which have been identified and listed as Stateless. Accordingly, as a way of demonstrating how the Function works, it has been used to identify the Pemba who live in Kenya as stateless people, and then suggest integration of them into the Neighboring Giriama or Rabai Community based on displayed intersecting characteristics. In operationalizing the Robust Nonparametric Kernel Discriminant Function, data from the Kenya National Bureau of Statistics (KNBS) obtained from the 2009 Kenya Population and Housing Census and a survey report on Pemba Community conducted in 2015 was applied to the study. Various characteristics associated with the listed Tribes/Ethnic Communities such as Education Level, Religion, Housing Building Materials (Housing Materials for the Floor, Walls and Roof), Waste Disposal, Source of Water and Employment Status, were considered. From the Theoretical developments and Empirical demonstrations, the findings from this study indicate that, the developed Nonparametric Discriminant Function provides a good classification method for classifying Stateless Communities. This is because they exhibit lower Misclassification Rates compared to the existing Parametric Methods. Use of the Kernel Discriminant Function is therefore recommended in classifying Stateless Persons. The study further recommends to the Government of Kenya to integrate the Pemba into either Giriama or Rabai communities and recognize them as Kenyan Citizens. Being that the methods developed and used herein are somewhat Global, results from this study respond to a major push by United Nations Human Commissioner for Refugees to "map" the size of Stateless Populations and their Demographic Profiles and respective causes, potential solutions and associated Human Rights Situations. By classifying/associating Stateless Communities to a particular Local, yet already existing and properly defined/known

Community that is recognized, a way of integrating them is one of the potential solutions, which then feeds into the greater Global Agenda regarding ending Statelessness across the world. This will help in making service delivery to such people without discrimination and go a long way in restoring their dignity.

CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

Nationality serves as the linkage between a Citizen and the International Systems through respective Domestic Laws. Nationality, traces its roots to the History of Human race with people having a sense of belonging to a Nation/Country. Hence, the Nationality to which an individual belongs guarantees him/her defined rights. Although, every person can have a right to Nationality, the same has not been experienced by every individual in our world today. This has created a situation where some individuals have been Stateless, (Milbrandt, 2011).

Statelessness exists in every region of the World, but remains a largely "hidden" problem without Government's recognition, (Manly, 2012). According to International Laws, a Stateless Person is someone not considered as a National by any state under the operation of its law. Statelessness is therefore the absence of any Nationality, which is the legal relationship or bond between the Citizen and his/her state based on social facts of attachment. Statelessness is a Global anomaly and many persons who are Stateless have never crossed an International Borders (Manly and Van Waas, 2014), (Sutton, 2018).

Two United Nations Conventions established the International Legal Framework for 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 a Stateless person and also provides Minimum Standards of treatment for Stateless Persons. The 1961 Convention on Reduction of Statelessness sets out Guidelines for prevention of Statelessness.

In Kenya, the Legal Framework relating to Citizenship is governed by the following Instruments: The Constitution of Kenya (2010), the Kenya Citizenship and Immigration Act (2011), the Kenya Citizens and Foreign Nationals Service Act (2011), the Kenya Citizenship and Immigration Regulations (2012) and various relevant Government Circulars. Chapter Three of the Constitution of Kenya (2010), lays out the different ways in which a person can become a Kenyan citizen. Section 15 (1) of the Kenya Citizenship and Immigration Act (2011) defines a Stateless Person as one who does not have an enforceable claim to the Citizenship of any recognized State and has been living in Kenya for a continuous period since 12th December, 1963. This definition only recognizes persons who were in Kenya at independence in 1963 and their descendants.

Section 16 (1) of the Act further states that a person who voluntarily migrated to Kenya before 12th December, 1963 and has been continuously living in Kenya [since then], shall be deemed to have been lawfully a resident. Such persons may be eligible for registration as citizens of Kenya, provided the person does not hold a passport or an Identity Document of any other Country, has adequate knowledge of Swahili or a Local Dialect, has not been convicted of an offense and sentenced to imprisonment for a term of three years or longer, intends upon registration as a Citizen to continue residing in Kenya permanently or to maintain a close and continuing association with Kenya, and the person understands rights and duties of a Kenyan Citizen.

Section 17 (1) of the Kenya Citizenship and Immigration Act (2011) provides that a person who has attained the age of eighteen years and whose parents are, or were, eligible to be registered as a Citizen either as Stateless or Migrants, may upon application be registered as citizens provided they fulfill some criteria. Descendants of Stateless Persons and Migrants need to satisfy that: there is sufficient proof that their parents are Stateless or Migrants; the person must have been born in Kenya, be continually residing in Kenya since birth and fulfill similar criteria as those required of Migrant Persons above.

Regulation No. 10 of the Kenya Citizenship and Immigration Regulations (2012) provides an application procedure for citizenship by registration of the Act under Sections 15 - 17 of the Kenya Citizenship and Immigration Act, (2011). The same Sections of the Act provide for definitions of and procedures for application into Citizenship for Stateless Persons, Migrants and their Descendants. These procedures have, however not been operationalized through a Gazette Notice, perhaps because there is no proper reference document on how it should be done.

Based on all the mentioned Legal Frameworks, the following are common causes of Statelessness: lack of Birth Registration and Certificates; birth to Stateless Parents; political change and transfer of territory, which may alter the Nationality Status of Citizens of the former State(s); administrative oversights, procedural problems, conflicts of law between two countries, or destruction of official records; alteration of Nationality during marriage or the dissolution of marriage between couples from different countries; targeted discrimination against minorities; laws restricting acquisition of Citizenship; laws restricting the rights of women to pass on Nationality to their children; laws relating to children born out of wedlock and during transit and loss or relinquishment of Nationality without first acquiring another, (Sutton, 2018).

At the end of 2019, the UN High Commission for Refugees counted 4.2 Million Stateless Persons worldwide, but estimated that the actual number may be over 10 Million due to under-reporting. At the Global Level, Stateless Persons are mainly found in Myanmar, Bangladesh, India, Indonesia, Malaysia, Kuwait, Thailand, Iraq and the Dominican Republic. Latvia and Estonia are some of the countries in Europe that host stateless persons. Closer home, many Stateless Persons are found in Burkina Faso, Mali, Ghana, and Cote d'Ivoire.

Kenya has a few groups who remain in protracted situations of Statelessness. These include the Pemba, people of Burundi and Rwanda descents, and children born in Kenya to British Citizens after 1983, (Muimi, 2021). The Galjeel, a Kenya Somali minority group, were stripped of their Nationality in the 1980s. Frequently, Stateless 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 Nationalities 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, it is estimated to be 18,500 after registration of the Makonde, (Abuya, 2010). 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 of children. This could result in potentially increasing the number of children that become Stateless or those who are stateless remain in that state 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 International Human Rights Community. The grounds thereof are based on Kenya's National Laws being inconsistent with her International Human Rights Obligations. In order to make an adequate assessment of its laws, it should be noted that the causes of Statelessness in Kenya can be divided into two broad categories, namely, Administrative and Legal, which illustrate the gap between Law and Practice.

The Administrative causes of Statelessness in Kenya such as the faulty operation or under-regulated nature of her Administrative Practices concerning Citizenship puts individuals, especially children, at risk of becoming Stateless, (UNHCR, 2014b). This is evinced in 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 Certificates causing additional travel costs and prolonged intimidating processes.

In Kenya, the known groups of refugees are the Galjeel and Pemba (Bosire, 2017). This was the case for Stateless Persons and those of undetermined Nationalities during the 2009 Kenya Population and Housing Census. It did not specifically categorize resident persons of unknown Nationalities in Kenya at that time. This was because the options on Kenyan Tribe and Nationality were grouped under one question and there was no provision for Stateless Code. All those who did not report as belonging to either a Kenyan Tribe or a specified Nationality were included in the *"other"* Category. It is therefore clear that Statelessness was not provided as a response option for those who were not Nationals of any state. However, some studies by the United Nation High Commission for Refugees estimates the Stateless Population in Kenya to be between 18,500 and 20,000, (Sutton, 2018).

Despite the attempts to improve on 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 to help in specifying each Stateless Community. Further, it established a population of only 6,272 as belong to a

Stateless Community. It is suspected that a majority of this group would hide their identities for fear of an imagined victimization.

The Global Action Plan include actions to resolve existing situations of Statelessness; prevent new cases from emerging and better identity 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 is publicly available for 150 States and that qualitative analysis on this group is publicly available for at least 120 States, (UNHCR, 2014b). In Kenya, a Stateless Household is one which has at least one person with family links to a Stateless Community. The link could be a person's parent(s) or grandparent(s) who migrated to Kenya for various reasons.

This study focuses on the Stateless Persons in Kenya and narrows down to the Pemba community who are estimated to have a population of about 4,000 in Kenya, (KNBS, 2009). It focuses on the Legal, Social and Economic Status of the Pemba people in the Coastal areas of Kenya. The first Pemba people arrived in Kenya in the 1930s and many settled among host communities, (Cole, 2019), (Manby, 2018). They came to Kenya from Pemba Island in two waves in the mid- 1930s and in the 1960s, as a result of the Zanzibar Revolt of 1964. Between 1935 and 1940 the first arrivals were seeking better livelihood opportunities. This community lives mainly in Kilifi and Kwale Counties, with a few living in Lamu County.

Persons born in Kenya with one parent from Pemba and one from Kenya should qualify for Kenyan Citizenship. Persons who have lived permanently in Kenya for at least seven years can become naturalized Citizens. Citizenship is also provided for if a person has been married to a Kenyan Citizen for a period of at least seven years. One key document accepted by the Government of Kenya for one to conduct day to day basic life activities is the National Identification (ID) Card. It is not only an Identification Document but also provides the basis on which residents are granted basic services.

Most Rights and Services are not limited to Citizens, however, participation in political activities like voting, is reserved for Citizens together with deliberate social welfare for groups like the youth, women and the elderly are reserved. This not withstanding, Stateless Persons cannot access some of the minimum Basic Services, (Abuya, 2010). Without any Nationality, Stateless lack the privileges and protections that citizens enjoy. While the Coronavirus (COVID-19) Global Pandemic posed certain risks to everyone, Stateless Persons stood a higher risk, (UNCR, 2020). First, they were more likely to have underlying health conditions and live in conditions that put them at heightened risk of infection. Second, they may not have had access to testing or treatment on equal basis with Nationals or foreign Nationals who are legally staying in the country. Third, they may have feared coming forward for testing or treatment because of their legal status, which put them at risk of detention or even deportation. Protecting everyone in the territory, irrespective of their Legal Status, is critical to any sound public health strategy. With Vaccines being available, some Stateless Persons and those at risk of Statelessness may not have been included in National Immunization Programs, (UNCR, 2020).

A considerable number of Stateless Persons have been in prolonged pre-removal detention as they are not considered legally resident and there is no Country to deport them to. However, a number of European countries have started to release asylum-seekers from detention and not to place additional people, including new arrivals, in closed facilities. This includes, for example, Austria, Belgium, Luxembourg, Spain, Switzerland and the United Kingdom. Similar measures have been taken in Senegal, Cameroon, Cote d'Ivoire and Burkina Faso, (UNCR, 2020).

Stateless Persons often already live on the margins of society, and lack of Legal Identity Documentation exacerbates their lack of access to social services. They may live in sub-standard, crowded housing with inadequate sanitation that compounds the risk of serious outbreak. They could not always adhere to public health protocols such as Self-Isolation and Social Distancing, making them more vulnerable to contracting the Virus. In Sudan, food distribution to vulnerable families is carried out without families needing to show individual documentation or a national identification number, (UNCR, 2020).

In Kenya from 1915 until 1947, ID cards were issued to all men in the country, including Europeans. From 1947 to 1978, ID cards were issued only to men of African descent. In 1978 an amendment to the Registration of Persons Act provided that ID cards were to be issued only to Kenyan Citizens who had attained 16 years. In 1980, this legislation was further amended to raise the age of majority to 18 and to include women. People with a foreign background residing in Kenya at that time were henceforth issued with Alien Cards, which limited their access to Government Services and Employment. Some Pemba were issued with Refugee Certificates in September 2009, whose validity ended in August 2016, (Manby, 2018). Although some Pemba were issued with IDs in Kenya, most of those 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 people 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.

Majority of the Pemba people are farmers and/or fishermen, many have very little education and poverty is prevalent. Moreover, the State places importance on the ID card as a form of proof of citizenship. Some of these people who claim to be eligible for Kenyan Nationality and are older than 18 years, do not have ID cards because they are not considered eligible. Lack of such a document has a circular effect of rendering a person at risk of Statelessness if there is lack of recognition that a person is a citizen. Access to some basic rights and services such as acquisition of birth certificates has been hard.

Access to 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 require an ID card. Stateless persons find it hard to participate in social and economic affairs such as owning property like land and even mobile phones. 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 advantage of Banking Services.

The Government of Kenya has been offering cash transfers to a number of vulnerable groups in Kenya. Such funds include Cash transfers for orphans and vulnerable children; older persons cash transfers and persons with severe disabilities. Some are given as loan facilities while others are grants. For example women enterprise fund (WEF) has both loans and grant components. Uwezo fund is given to women, youth and persons with disability in the form of loans and grants. Youth enterprise fund is also given in the form of loans and grants.

Further, there is National Government Affirmative Action Fund (NGAAF) which is flagship project for vision 2030 under the Social Pilar. The fund is meant to address the plight of vulnerable groups through enhanced access to financial facilities for social-economic empowerment among women, youth and PWDs, needy children and the elderly in Kenya, (KNBS, 2021). All these facilities are not available to the corresponding Pemba groups and their freedom of movement is also limited.

The descendants of Stateless Persons have a high risk of not obtaining IDs upon attaining the adulthood age because they are required to produce a copy of the parents' ID cards during the application for their own cards. Those who have been unable to apply for a Birth Certificate within the first 6 Months of birth are locked out of the process for good since they cannot provide any identity documents to support the procedure. Many have turned to "buying" parents for their children to ensure that they are registered for the Primary School Level Examination. Because of desperation some Pemba 'give up' their children to Kenyans in order for them to be registered using Kenyan Names and be able to access Education.

The older generation of the Pemba have developed ***coping mechanisms*** to acquire ID cards and other vital documents such as using Pseudo-Names or a Kenyan Tribe. In Kilifi, many families have resorted to registering fraudulent names of fathers as a coping mechanism to enable children to enroll in school. The children may in turn be forced to take a Pseudo-Name when they apply for an ID card once they attain the age of majority. This may be one reason why many of the Pemba did not identify with a particular culture or tradition. In 2007 there were government attempts to address issues of the Pemba Community and many had their bio-data collected but the exercise did not bear fruits.

Attempts to register with the Tanzanian Authorities to obtain identification documentation as a last resort were rejected on the grounds that they have no links to Tanzania. The Pemba speak Swahili fluently, which is a National Language in Kenya and also the Local Language in Pemba. Some also speak Local Tribal Languages, which is an indication of some Integration into the Communities where they live. Among the Stateless Communities, Nubians were the first to be given Kenyan Citizenship. This was followed by the Makonde Community in 2017 and later on the Shona Community. However, Pemba and Galjeel people are still waiting for the same.

The study therefore looks into how the Pemba Community can be integrated into some of the Local Communities. The focus is on the Pemba Community because it has second set of data arising from a survey that was conducted in 2015. Thus, it is important to fully understand characteristics of the Pemba Community and find out if there are any similarities with 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 were to be absorbed.

To achieve this, a Nonparametric Kernel Discrimination Function is used for 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-classifications rates are computed and compared with other existing Classification Models.

1.1.1 Statelessness

Stateless persons and those at risk of being in this situation have been an ongoing issue both at a domestic level as well as internationally. Individuals residing in a number of countries are being denied the privileges and rights given to the rest of the population because they have no documents to prove their citizenship, Kerwin and Warren (2019). In many African Countries, Stateless people especially children, face discriminatory and arbitrary Nationality laws as a result of which they are not registered and granted Citizenship in their Country of birth or where they are found or undocumented.

Thus, they continue to be Stateless and will not be able to register their own children once they become parents. As a result, this creates an issue of Trans-Generational Statelessness which will continue indefinitely and as such, requires attention and action both at Domestic and International Levels as a matter of urgency. While laws have been enacted with the aim to protect Stateless People from the risk of continuously being and becoming Stateless, lack of Guidelines in the implementation thereof creates difficulty for children to acquire a Nationality.

States in this regard have the responsibility to create mechanisms of facilitating the implementation of laws especially when dealing with vulnerable groups such as Stateless Children. In this regard, the current study seeks to offer a solution Statelessness, which is not only a problem in Kenya, but Globally.

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 of children. This could result in potentially increasing the number of children that become stateless and those who are Stateless will remain the same way.

Kenya has not ratified the 1954 Convention relating to the Status of Stateless Persons and the 1961 Convention on Reduction of Statelessness. Nevertheless, the discriminatory Nationality Laws and Administration thereof, have repeatedly been brought to the attention of International Human Rights Groups. The grounds thereof are based on Kenya's National Laws being inconsistent with Kenya's International Human Rights Obligations. For purposes of adequate assessment on 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 then 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, (UNHCR, 2014b). This is evinced in 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 Certificates and various additional documentation which require repeated trips to various government buildings causing additional travel costs and a prolonged intimidating process.

1.1.2 Causes of Statelessness

Statelessness can be caused by numerous factors. Some of these are of a Legal Technical nature, where statelessness is caused by gaps in Nationality Laws or conflicts between Nationality Laws. States determine their own nationality laws, within certain limited restrictions imposed by International Human Rights Law. The two main Legal Principles Governing States' grant of Nationality at birth are *jus sanguinis* (Citizenship by Descent) and *jus soli* (Citizenship by Birth in the Territory).

Conflicts in these laws are one of the several types of conflicts of law situations that can render a child Stateless. For example, a child born in the territory of a *jus sanguinis* State to parents with Nationality of a *jus soli* State would encounter problems obtaining any Nationality if the National Legislation of the two States relevant here does not contain provisions that would allow such a child to obtain Citizenship. Statelessness can also occur later in life. Some Legal Systems provide for mechanisms of automatic loss of Nationality, for example after a long absence from the territory. Some States require that a person renounce his or her previous nationality before acquiring the Nationality of that State. Withdrawal of Nationality can also lead to Statelessness if there is no adequate safeguard in place to prevent Statelessness.

Another major cause of Statelessness relates to the dissolution and separation of States, disputes about borders, transfer of territory between States, and the creation of new States. In the period of decolonization, groups of persons may have been left out of the initial body of Citizens under the Nationality Legislation of the newly independent State. In Europe, many people were left Stateless after dissolution of the Soviet Union and the Socialist Federal Republic of Yugoslavia.

In addition to the aforementioned causes of Statelessness, discrimination in Nationality Law or in Practice against certain parts of the population and arbitrary deprivation of Nationality contribute significantly to the creation or perpetuation of Statelessness. Based on, for example, Ethnicity or Religious Beliefs, a certain group within a State or populations living across multiple States are sometimes denied or deprived of Nationality. Examples of such populations are the Rohingya in Myanmar, the Bidoon in the Arab Gulf States, and parts of the Roma population in Europe.

Today, 27 States still discriminate against women in their laws with regard to transmission of nationality to children, the majority of which can be found in Africa, Asia and the Middle East. Further, laws that discriminate against children born out of wedlock, for example by making it more difficult for them to acquire their father's nationality, can also contribute to Statelessness, (UNHCR, 2014a).

1.1.3 Consequences of Statelessness

Most Stateless persons encounter many difficulties in every aspect of daily life. Often, Stateless persons do not enjoy their basic human rights. Even though the enjoyment of fundamental human rights is not formally dependent on Citizenship status, many States extend human rights protection to their nationals only or to persons who reside lawfully in the country, which is not always the case of stateless persons. Stateless persons may face obstacles accessing education or health care services, entering the labour market, traveling abroad, or owning land or other property. Stateless persons may not be able to register the birth of their child, obtain an identity document, open a bank account, inherit wealth, or get legally married.

Being Socially and Economically excluded, Stateless persons are vulnerable to abuse and destitution, and many Stateless populations belong to the most marginalized and vulnerable groups worldwide. Also, Stateless Persons may be detained for prolonged or repeated periods because they have no identity documents or because they are considered to be irregularly in the country, yet there is no country to which they can be returned.

1.1.4 Review of International and Regional Instruments Relevant to Nationality and Statelessness

As established in various Reviews within Chapter 1, the right to acquire a Nationality is a fundamental right under International Law. Therefore, the absence thereof creates the potential for states to abuse their power by discriminating against Stateless Persons. While many states fail to realize the importance of a right to a Nationality and its responsibilities coupled thereto, Citizenship is an ever-present issue and often a major obstacle. This is because recognition of nationality serves as a key to a wide range of other rights, such as health care, education, employment and equality before the law, (Bloch and Donà, 2019). In this respect, it is clear that persons and especially children who are stateless are some of the most vulnerable groups in the world.

Each state has the sovereign responsibility to determine under national law who are its citizens and who are not or who can qualify or who cannot qualify to be a citizen. However, that role is subject to international principles. While international law principles may be ratified by states, there are no international enforcement mechanisms in place should such states fail to adhere to the said principles. It is required that, once a state has ratified an international treaty, its principles are incorporated into the domestic law of the state by which failure to adhere to such principles would amount to domestic sanctions.

There are a number of international and regional instruments that affirm the right to nationality or the right to acquire a Nationality which are discussed below in order to give understanding of such instruments and its relevance in the jurisdictions chosen.

The right to Nationality is enshrined in the Universal Declaration of Human Rights (UDHR) and many other International Instruments and Laws. In particular, the Human Rights Law, has increasingly recognized an individual's right to a Nationality. The right to Nationality generally requires appropriate states to grant it to individuals who would otherwise be Stateless. These International Laws are essential in the protection and prevention of Statelessness among people and thus, important in attending to the purpose of this study.

Universal Declaration of Human Rights (UDHR)

While no international consensus has been reached on the definition or classification of statelessness, in order to alleviate the suffering of persons at risk of statelessness, multiple legal instruments have been adopted at international and regional levels. These efforts to formally protect stateless persons can be traced back to 1948 at the promulgation of the UDHR, (Abuya, 2010). It was the first legal document to set out the fundamental human rights to be universally protected.

An example that lays this foundation is Article 15 of the UDHR which provides that *"Everyone has the Right to a Nationality"* and that "No one shall be arbitrarily deprived of his nationality, nor denied the right to change his nationality", (Assembly, 2014). The inclusion of the Right to Nationality in Article 15 of the UDHR, like the UDHR as a whole, was motivated by the impulse to respond to the atrocities committed during the Second World War, among them the mass denationalization and huge population movements, (Bloch and Donà, 2019).

There were hundreds of thousands of Jews who survived the Nazi-perpetrated genocide and fled their home countries, while millions of ethnic Germans were expelled from eastern European states, and millions of Poles, Ukrainians, Byelorussians and other minority populations of the Soviet Union either were forcibly expelled or fled for their safety, (Bloch and Donà, 2019).

In 2005, the Commission passed a resolution reaffirming Article 15 of the UDHR, emphasizing that the right to Nationality of every human person is a fundamental human right, (Onu and John, 2016). The United Nations Special Rapporteur on the situation of human rights in the Democratic Republic of Congo (DRC) has also based findings on the right to Nationality.

For example, in 1996, the Special Rapporteur found that the government of DRC had violated the Banyarwanda and Banyamulegue people's right to Nationality guaranteed in the UDHR as well as the customary International Law prohibition against Statelessness, (Garreton, 1997). These changes in law provided a great foundation to the increasingly recognized and protected right to Nationality in which other International Instruments have based their Nationality Laws upon.

The Statelessness Conventions

The 1954 Convention relating to the Status of Stateless Persons (1954 Convention) and the 1961 Convention on the Reduction of Statelessness (1961 Convention) specifically addresses the issue of Stateless Persons. While the 1954 Convention establishes a definition of a Stateless Person, (Van Waas-Hayward, 2008). and a set of minimum rights which protect those who are currently stateless, the 1961 Convention is grounded on the aim to prevent statelessness. Both the 1954 Convention and the 1961 Convention have a different purpose but combined they form part of an important contribution to eradicate Statelessness.

1954 Convention Relating to the Status of Stateless Persons

The 1954 Convention provides for the identification, documentation and protection of the rights of stateless persons. It additionally provides an internationally recognised status for stateless people and a framework for States to protect Stateless people, (Southwick and Lynch, 2013).

The 1954 Convention confirms that stateless persons retain fundamental rights and freedoms without discrimination These fundamental rights include free access to courts, primary education, public relief at par with what the State's Nationals receive and property rights, access to employment and housing at least as favorable as those afforded foreign persons and more specifically with regards to children, safety and physical well-being of such children who are stateless or at risk of becoming stateless, (Southwick and Lynch, 2013).

The definition of a Stateless Person is established under Article 1 of the 1954 Convention which provides that a Stateless Person is someone not considered as a National by a State under the operation of law, (Assembly, 2014). In order to establish whether or not a person is a national under the operation of a state's law, it requires a careful analysis of how the State applies its National Laws in practice, (Manly and Van Waas, 2014).

If and when persons satisfy the definition of a stateless person, such persons are entitled to certain rights and must comply with certain duties contained in the 1954 Convention. It is important to note that the 1954 Convention does not cover the so called *de facto* Stateless Persons for whom no universally accepted definition exists in International Law. However, despite the fact that *de facto* Stateless persons are not covered by the 1954 Convention definition, such persons are entitled to protection under International Human Rights Law. For example, Stateless Refugees are covered by the 1951 Convention relating to the Status of Refugees and should be treated in accordance with International Refugee Law, (Assembly, 2014).

The 1954 Convention is based on a core principle that no Stateless person should be treated worse than any foreigner who possesses a Nationality. With the Human Rights System being grounded on the concept of universality, lack of Nationality should not act as an automatic barrier to enjoyment of its guarantees, (Weissbrodt, 2000).

However, the reality thereof is significantly different. This is due to the fact that there are practical difficulties of accessing rights without identity documents or proof of lawful residence in a country as well as certain key rights are reserved explicitly for nationals. For example, the right of political participation, the right to work and the right of entry in a country, (Manly and Van Waas, 2014). Additionally, the underlying principle of non-discrimination in International Human Rights Law does not preclude any distinction between Citizens and others, (Assembly, 2014). Instead of preventing such distinction, differentiation is permissible so long as it furthers a legitimate objective and sits within the bounds of proportionality, (Manly and Van Waas, 2014).

In considering the predicament faced by stateless persons, the Convention stipulates that they must be treated like nationals of the State in respect of certain rights such as freedom of religion or elementary education, (Assembly, 2014). It must be stressed that the Convention pursues a nuanced approach in which it specifies that some guarantees apply to all stateless persons while others are reserved to stateless persons who are lawfully present or lawfully staying in the territory. Thus, the 1954 Convention echoes human rights standards contained in other international instruments and provides guidance on how such standards are to be implemented for Stateless Persons.

In this regard, Stateless Persons have the duty to obey the Laws and Regulations of the Country in which they find themselves in as set out in Article 2.73. The question here is, how stateless persons who have been living in some of the host countries countries can be integrated/ classified within the local communities so that their numbers can reduce as proposed by UNHCR.

1961 Convention on the Reduction of Statelessness (1961 Convention)

Subsequent to the enactment of the 1954 Convention, the 1961 Convention was adopted in 1961 and entered into force in 1975. It compliments the 1954 convention and was as a result of international negotiations on how to avoid the incidence of statelessness. It focuses on avoiding statelessness from birth and prevents the creation of statelessness as a result of loss, deprivation or renunciation of a nationality, (LawyersforHumanRights, 2014). The 1961 Convention contains rules implemented through Nationality laws to ensure that everyone enjoys the right to Nationality in practice, (Persons, 2012).

There are four main areas in which the 1961 Convention provides concrete and detailed safeguards to be implemented by States in order to prevent and reduce statelessness. The 1961 Convention does, however, allow for limited but significant exceptions to these obligations and prohibitions. This is due to the fact that its provisions have been supplemented by the subsequent implementation of international human rights law in relation to nationality, (Manly and Van Waas, 2014).

1.1.5 The General Classification Problem

It is a well known fact that every unit of a population possesses some characteristics similar to or different from other members of the group in which its membership is established. Hence the application of discriminant analysis is generally motivated by the need to reclassify some individuals who by one reason or the other belonged to the wrong group(s) and need to be placed in the correct group. The need for discrimination arises if there are some individuals in a group whose characteristics are significantly different from the general characteristics of members forming the greater proportion of number in the group.

Patterns are considered to be the means by which the world can be interpreted. Based on this idea, people are able to read a book and recognize every character or image included in the pages. This ability is based on knowledge gained by experience in reading these same characters or seeing similar pictures. Using similar rules (or experience) people are able to discriminate between different colours, sizes, faces, etc.

This concept motivated Scientist to develop methods of solving other types of problems, such as discrimination between Benign and Malignant Tumors, (Mangasarian, 1965) the detection of fraudulent transactions, (Brause et al., 1999) and discrimination between bad and good payers, e.g. (Thomas et al., 2017). All these problems are set under the general label of classification.

Specifically, in classification the aim is to assign observations into a number of prespecified classes so that the objects in the same class are similar to one another, (Garden, 1981). After learning these patterns a model is used to classify new examples. The process of classification from a model development aspect consists of several steps: data collection, data preprocessing, feature selection, classifier development, and assessment of the results. Data collection is very important because data quality affects the quality of the results. The GIGO (Garbage-In-Garbage-Out) principle characterizes classification problems because the final results depend on the data used as inputs to the process.

Therefore before using the data it is important to apply some preprocessing actions such as data transformation, sampling or feature selection. The latter action is used in making the classifier more flexible and possibly more accurate when applied to different data than the data used in the development. When assessing the results, it becomes important to use the most appropriate criterion depending on the nature of the problem as some measurements are less accurate under some data conditions such as imbalanced class sizes. The whole process is iterative partially or overall, e.g. feature selection can be repeated several times until the optimal subset is found and also some of the steps can be missed.

Classification is relevant to a large range of problems such as cell tissue analysis, (Sun and Xiong, 2003) , heart disease, marketing, and diabetes, (Adams and Hand, 1999). An area that has received much attention during the last three decades is credit scoring. In credit scoring, lenders use data from previous borrowers in order to discriminate between customers that might go bad (miss a number of consecutive payments) and good (who will not). This approach is used for a range of different products such as credit cards, auto loans, personal loans, small business loans and mortgages.

1.1.6 Classification Problems involving Stateless People

In Social, Economic and Industrial Problems, we are often confronted with the task of classifying an individual into one of two groups on the basis of a number of test scores. Therefore, the problem of classification will arise when an investigator makes a number of measurements on an individual and wishes to classify the individual into one of several categories or population groups on the basis of these measurements, (Flury, 2013). 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.

A similar situation arises in connection with college entrance examinations. Again, on the basis of a number of test scores, the admission or rejection of a student has to be decided. In all such 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 represents 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.

Often, some statistical data from past experience are available which can be utilized in making the classification. Suppose that from past experience we have the test scores of N_1 individuals who are known to belong to population P_1, and also the test scores of N_1 individuals who are known to belong to population P_2. These data will be utilized in classifying a new individual on the basis of his test scores.

Similarly, just like in other sectors as discussed, classification can be very vital in solving the problem of statelessness in a country. Related to the focus on National Policy, as proposed by UNHCR and an array of partners which advocate for countries to 'map' the size of Stateless populations and their demographic profile, as well as causes and find a potential solutions to this statelessness. Any method that can propose integration of stateless persons in the existing Communities will be aligned in the UNHCR Goals and Objectives.

1.1.7 Discriminant Function as a Tool for Solving Classification Problems

Discriminant Analysis (DA) is one of the popular multivariate methods which has a long history. DA is a classification problem that consists of assigning or classifying an individual or object to one of several known or unknown alternative classes (or groups) on the basis of many measurements on the individuals or objects, or cases. For instance, the Bayes Discriminant Rule provides that a point x can be classified into one of the groups j such that

$$x \text{ is allocated to group } j_0 \text{ if } j_0 = arg \max_{j \in 1,\ldots,v} \pi_j f_j(x) \tag{1.1}$$

where π_j is the prior probability of drawing from density f_j.

The goal of discriminant analysis is given a data set with two or more than two classes (or groups), say, what is the best feature or feature set either linear or non-linear to discriminate between the classes and maximize average class separation or equivalently minimize the probability of misclassification.

The Classical Approach to Classification

The classification problem in Discriminant Analysis (DA), which involves assigning (classifying) entities (observations) to exactly one of several well-defined mutually exclusive groups or classes, based on their characteristics on a set of relevant attributes, is important in almost any field of Applied Sciences.

Many different approaches have been proposed for solving the Classification Problem in DA. The Classical Approach to Classification is to first estimate the probability (density) functions and then derive the Classification Rule that minimizes either the probability of misclassification or the expected Misclassification Cost. A second approach is to estimate the posterior probabilities of group membership directly, and use a classification rule that weighs these probabilities by the appropriate misclassification costs.

A third approach is to specify a particular form of Classification Function, and then determine the parameter values of this function that optimize some accuracy criterion-that is, some measure of Classification Accuracy in the training sample. Generally, Application of Classification has been employed in various fields such as Prediction and Forecasting Tasks, Diagnosis Tasks, and Pattern Recognition.

1.2 Statement of the Problem

There are many people who are Stateless or at Risk of Statelessness. These include descendants of people who migrated from other places many years ago, and whose children are unable to establish rights that ought to be derived from their parents. Such people include the Comorian Migrants from Zanzibar, Galjeel and Pemba. In most cases, Stateless Persons are not only undocumented but also overlooked and not included in National Administrative Registers or even Databases.

Many Stateless Persons of undetermined Nationality are normally counted during Censuses but not given a specific Ethnic Classification. This leads to inaccuracies since the Stateless Persons cannot be recorded as Nationals of their original countries or Refugees. This challenge is currently of a Global concern, (Milbrandt, 2011; UNHCR, 2014a). Accordingly, there is eminent need to have an approach that can be used in identifying Stateless Persons from a given population, (using a robust tool, based on certain characteristics, to "discriminate" such persons from a population and correctly classifying them) with an intention of integrating them with the existing Local Neighbouring ethic Groups, (UNHCR, 2017).

To do this, we need a statistical approach that would lead us to the final results. Basically, there are two possible approaches that could be applied here. They are Parametric and Nonparametric Approaches. The existing literature shows that in the efforts to do Discrimination and implied Classification, many previous studies have used Parametric Approaches like the Linear Discriminant and Quadratic Discriminant Functions. However, these Parametric Approaches rely

on Normality and Linearity Assumptions, Park and Park (2005), yet in the real world, the problem of Statelessness defies these Assumptions. Therefore, a procedure that can stand the test of time and be admissible to the affected groups is required.
An alternative approach to apply is Non-parametric discriminant methods which are based on Nonparametric group-specific probability densities. For this approach, either a kernel or the k-nearest-neighbor method can be used to generate a non-parametric density estimate in each group and to produce a classification criterion. In this case, the performance of a discriminant criterion was evaluated by estimating probabilities of misclassification of further observations, (Fernandez, 2002).
Models have been developed extensively in the area of parametric methods but just a few on non-parametric ones. For the few Nonparametric methods in the literature, k-nearest neigbour has been exploited to some extent. In this study, we proposed the use of Nonparametric kernel-based method. On the other hand, the practical solution of stateless persons has not been exploited using Nonparametric approach.

It is for the above reasons that this study proposed to use a Nonparametric approach with an appropriate Kernel function then apply data from past Censuses and Surveys through discriminant analysis by establishing a classification criterion to identify a Local Community whose characteristics are close to those of the Stateless Persons under consideration here for purposes of Integration.

1.3 Justification of the study

1.3.1 Justification to the Theory of Statistics

This study demonstrates that it is possible to replace the Parametric Functions within the general Bayes Discriminant Rule with an optimally chosen Kernel Function thereby yielding a Robust Nonparametric Discriminant Function. The performance of this developed Robust Nonparametric Discriminant Function is better when compared with Linear and Quadratic Discriminant Funtions. This is therefore a Milestone when used in context of Survey Sampling, treating the local communities as a parent population with heterogeneous characteristics from which a sample is identified and correctly classified.

Additionally, in the field of classification especially the use of Nonparametric discriminant function has not been explored much in literature especially in the application when classifying the stateless people. On the other hand, much has been discussed and application made in the parametric discriminating functions. The challenge with these functions is that they fail to capture the nonlinearity that occurs mainly in the real life datasets.
Further, most of the real life dataset do not assume the normality assumptions that are imposed by the parametric discriminating functions such as Linear and Quadratic. Therefore, this study will add in literature and for application purposes, another alternative discriminant classification method that can be used and relied upon in situations where parametric functions fail. Lastly, the findings of these study will add to a pool of literature where other researchers and scholars can make references in relation to classification of Stateless Communities since this area has not been fully exploited.

1.3.2 Justification to Users of Statistics and Other Stakeholders

An attempt was made to include some important variables in the 2019 Kenya Population and Housing Census to establish the population and characteristics of stateless persons in Kenya. This was a lesson from the previous Census and so a way of capturing Stateless persons was explored and a decision was made to include statelessness as an option under ***Ethnicity/Nationality***. However, it still lumped the stateless groups without specifically breaking down each of them. Furthermore, the 2019 Census results showed low numbers than expected because some stateless persons opted not to declare their stateless situation possibly for fear of victimization by the authorities.

The Global Sustainable Development Goals (SDGs) or the Agenda 2030 is founded on the principle of "Leave No One Behind (LNOB)". It represents the unequivocal commitment of all United Nation member states for example, to end discrimination and exclusion, and reduce the inequalities and vulnerabilities that leave some people behind. Failure to achieve the above would undermine the potential of individuals and of humanity as a whole. One of the causes of people being left behind is persistent forms of discrimination, including failure to recognize such vulnerable people, which leaves individuals, families and whole communities marginalized and excluded.

The spirit of LNOB compels us to focus on discrimination and inequalities (often multiple and intersecting) that undermine the people as holders of rights. In as much as SDGs stress on leaving no one behind, these groups of people may be left further behind if they are neither given citizenship nor fully assimilated into the surrounding communities. Without citizenship, stateless people have no legal protection and no right to vote, and they often lack access to education, employment, health care, registration of birth, marriage or death, and property rights. They continue encountering travel restrictions, social exclusion, and heightened vulnerability to sexual and physical violence, exploitation, trafficking of persons, forcible displacement, and other forms of abuses.

With the COVID-19 Global Pandemic, many Stateless Persons around the world could easily be missed out on vaccination especially in the early days when vaccines were discovered and made available by governments. They were excluded from vaccination programs either deliberately by defacto because they lacked proof of legal identity. While some countries expressly barred undocumented persons from getting vaccinated, in other contexts they were in principle eligible. For example, in Malaysia all persons present on the territory were urged to come forward for testing if they showed symptoms of COVID-19 and provided assurances that nobody would face detention or deportation, regardless of their Nationality Status or Legal Resident Status. In Portugal, all migrants and asylum seekers residing in the Country who who had pending applications were granted full access to the Country's Healthcare Services, (UNCR, 2020).

The documentation requirement in practice stemmed from the need to keep track of who had been vaccinated to invite them for subsequent inoculations and to track the safety of vaccines. Lack of documentation also makes it harder for the authorities to reach these populations as they typically do not appear in civil registers or national population registers hence invisible to the authorities. With majority of stateless Communities in Kenya and around the word remaining undisclosed, the findings of this study will be important to the Government of Kenya since it will guide on which local community, could be used to absorb or *"integrate"* Stateless Communities with. This will help in making service delivery easy to such people and go a long way in achieving the recommendation by the UNHCR.

1.4 Objectives of the Study

1.4.1 General Objective

To classify the Stateless Communities using a Robust Nonparametric Kernel Discriminant Function.

1.4.2 Specific Objectives

1. To develop a Nonparametric Discriminant Function from the Bayes Discrimination Rule using Kernel Discriminant Function.

2. To estimate the Classification Rates of the developed Nonparametric Kernel Discriminant Function as a measure of its Robustness.

3. To compare the developed Nonparametric Kernel Discriminant Function with the Linear Discriminant Function and the Quadratic Discriminant Function through a simulation study.

4. To apply the developed Nonparametric Kernel Discriminant Function in classifying stateless Communities in Kenya.

1.5 Organization of the Thesis

This Thesis has been organized as follows; the first chapter presents a background of the study, Statement of the problem, Objectives and a Justification. Chapter Two that follows, presents the Literature Review, with specific pointers to reviews of Parametric and Nonparametric Discriminant Functions, Classification Rates and Reviews of Performance Measures. Chapter Three has the Methodology used including the development of a Nonparametric Kernel Discriminant Function, the Misclassification Rates and the Algorithm of Classification for Kernel Discriminant Function and of Stateless Communities.

Chapter Four contains the Results and Discussions in which Simulation is done and an application of the developed Kernel Discriminant Function in the Classification of Stateless Pemba Community in Kenya is demonstrated. Finally Chapter Five gives Conclusions and the Recommendations resulting from the Chapters One up-to Four. In this final Chapter, an attempt has been made to answer the main question that was posed in the Problem Statement and implied in the General Objective of this Study.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

The problem of Data Classification is not new. It represents a Primary Goal in the fields of Pattern Recognition and Statistical Machine Learning. In it, Feature Extraction and Dimensionality Reduction is one of the commonly used preprocessing procedures for classification. Eigenvector-based Techniques, such as Principal Component Analysis (PCA) and Fisher's Discriminant Analysis (FDA), (Fisher, 1936), are well known techniques for Feature Extraction and Dimensionality Reduction.

Feature Extraction for classification differs significantly from Feature Extraction in describing data. For successful classification, Supervised Learning Techniques are preferred compared to Unsupervised ones. For instance, the Unsupervised Technique, PCA finds directions that have minimal reconstruction error by describing as much variance of the data as possible, (Mika et al., 1999). In some cases, it may discard dimensions that contain important discriminative information between classes. Futher, PCA are less interpretable and don't have real meaning since they are constructed as linear combinations of the initial variables. Unlike PCA, LDA seeks to find the projection directions along which the classes are best separated, (Yu and Yang, 2001).

When dealing with such problems, the main challenge that ought to be addressed is the matter of accuracy which may also be referred to as Robustness of the classifier. Accordingly, each time a classification function is to be used, matters of robustness in the classification function are cardinal. The current problem of this thesis is on the classification of stateless people, a problem whose nature does not allow for Trade-off between Robustness and any other statistical properties.

Herein, to ensure Robustness in the classification procedure, a Nonparametric Kernel Discriminant Function is used. The driving reason for the choice of a Nonparametric Kernel Discriminant Function is informed by the fact that Kernel-based approaches have been found to be a better choice whenever a non-linear classification model is needed, (Cochran, 2007). Once a Robust Kernel Discriminant Function is identified, it is then employed within the context of discriminant analysis to aid in classification of stateless people. Discriminant Analysis is concerned with the problem of classification. It arises when an investigator makes a number of measurements on an individual and wishes to classify the individual into one of several categories or population groups on the basis of these measurements, (Flury, 2013).

This chapter is organized into three sections, with the first part focusing on the review of both parametric and nonparametric functions. This is done in line with Objective 1. The second section looks at the previous work done in connection with a measure of robustness of an estimator so that it responds to the background details for Objective 2. The review of statelessness is covered in the next section and it responds to the study of Objectives 3 and 4. A summary of the chapter is given as a prelude to the discussion of the data gaps that eventually leads us to the main area of the study.

2.2 Theoretical Review

2.2.1 Review of Parametric and Nonparametric Discriminant Functions

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 analyst may wish to classify loan applicants as low or high credit risks on the basis of the elements of certain accounting statements, (Manly and Van Waas, 2014). (Lachenbruch, 1974) 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. (Johnson et al., 2014), 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.

(Usoro, 2006) viewed discriminant analysis as a multivariate statistical method which seeks to differentiate between two groups with respect to certain characteristics. According to him, the objective of discriminant analysis is to have a critical study in naturally or unnaturally occurring groups, and to distinguish between two or more predefined groups. He further stated that in most cases observations or individuals under careful examination may possess some properties which restrict their fundamental membership to two predefined groups of populations of which one group will have standard and similar properties amongst its members distinct from the other. In spite of the common similarities that exist in group membership, there may be some differentials based on certain characteristics which could make some members of a particular group or population different from other members. He went a head and applied discriminant analysis in classifying students on the basis of their academic performance. He 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.

(Friedman, 1989) observed that the formal purpose of classification or discriminant analysis is to assign objects to one of several, k, groups or classes based on a set of measurements $X = x_1, x_2, \ldots, x_k$ obtained from each object or observation. He further observed that classification techniques are also used informally to study the separability of labeled groups of observations in the measurement space. He also stated that in the formal setting, an object is assumed to be a member of one and only one class, and that an error is incurred if it is assigned to a different one.

(Flury, 2013) gave an example of discriminant analysis problem as follows: prospecting students applying for admission to college are given a battery of tests; the vector of scores is a set of measurements X. The prospective student may be a member of one population group consisting of those students who will successfully complete college training or rather have potentialities for successfully completing training, or he may be a member of the other group; those who will not complete the college course successfully. The problem is to classify a student for admission on the basis of his scores on the entrance examination.

(Erimafa et al., 2009) 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. (Thomas and Pascal, 2013) 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.

So far in the foregoing studies, the researchers relied upon the parametric discriminant methods. These methods are conceptually simple and would allow one to make generalization the results from a Sample to a Population and have been used in many application areas. However, their reliance on Linearity and Normality assumption limits their performance. Furthermore, as a linear method (Fisher Discriminant Function), it does not capture non-linearly clustered structures (Park and Park 2005).

Currently, a prevalent extension to the nonlinear problem is kernel discriminant analysis (KDA) (Roth and Steinhage, 1999), (Baudat and Anouar, 2000). The KDA first maps low dimensional data into a high-dimensional one, and subsequently projects high-dimensional data onto a low-dimensional one. It is able to recognize certain simple nonlinear relationships. However, KDA, in complex nonlinear structures, is not as effective as the Nonparametric method with a local classifier, such as k-nearest neighbor (KNN). Nonparametric discriminant analysis (NDA) relaxes the normality assumption of traditional LDA, (Fukunaga and Mantock, 1983). (Fukunaga and Mantock, 1983), Fukunaga et al., (1990), introduced the Non-parametric Discriminant Analysis (NDA) that has overcome the parametric form of FDA by extending the commonly used scatter matrices. As NDA relaxes the normality assumption of the LDA, it can deal with non-normal data distributions by incorporating data direction and boundary structure in its Within-Class and Between-Class Scatter Matrices, respectively.

The latter is computed based on all data points with respect to their k-nearest neighbors for each point, instead of relying on class means only as in LDA. This has made the scatter matrices generally of full rank, and therefore, the method can work well even for non-Gaussian data sets, as well as the ability to specify the number of desired extracted features. The NDA uses a weighting function for each data sample, to preserve classification structure by deemphasizing samples far from the classification boundary.

(Friedman, 1989) introduced Regularized Discriminant Analysis (RDA) proposing to add a regularizing parameter to within-class scatter matrix to minimize the generalization error. (Loog et al., 2001), introduced a weighted version of FDA aiming to downplay the roles of the class distributions that are farthest apart. (Li et al., 2013), provided new formulation of scatter matrices to extend the two-class NDA to multi-class cases. They also developed two more improved multi-class NDA-based algorithms (Nonparametric Sub-space Analysis (NSA) and Nonparametric Feature Analysis (NFA)) with each one having two complementary methods based on the principal space and the null space of the intra-class scatter matrix, respectively.

A combined model for NDA and support vector machines (SVM) was introduced by (Ksantini and Boufama, 2012). It aims to control the spread of data, while maximizing a relative margin separating data classes. This is done by incorporating the data spread information represented by the dominant normal directions to the decision boundary. However, these methods behave linearly with respect to class separation and fall into the problem of under fitting with high level of misclassification when data classes are not linearly separable. As real data distribution is generally nonlinear, the demand for efficient non-linear classifiers has been growing. Many researches have shown that kernel-based linear methods are computationally efficient, robust and stable for pattern analysis and classification, (Shawe-Taylor et al., 2004); (Melgani and Bruzzone, 2004); (Ksantini et al., 2011).

In these methods, the original data points are first mapped into a higher-dimensional feature space then, a linear method is used for classification. Such mapping is performed by a nonlinear function through a mathematical process called the *"kernel trick"* (Aizerman et al., 1964); (Boser et al., 1992). Hence, when using a kernel with any linear method, the originally linear operations are done in a reproducing Hilbert space, obtained through a nonlinear mapping.

Many existing linear algorithms have been extended to work with the Kernel Technique and some of those are Kernel Support Vector Machines (KSVM), (Boser et al., 1992) Kernel Principal Component Analysis (KPCA) , (Schölkopf et al., 1998), Kernel Fisher's Discriminant Analysis (KFDA or KDA), (Mika et al., 1999), and Kernel Canonical Correlation Analysis (KCCA), (Akaho, 2006). All these Kernel-based Algorithms are considered to be extensions of their linear versions to non-linear distributions. By applying the Kernel trick using some Kernel Function $k(x, y)$, the data is transferred from its original space X to an inner dot-product space K, where mapping features is implicitly performed.

Choosing the right kernel function, for instance, is crucial to the success of the linear model in the feature space (Howley and Madden, 2005); (Alizadeh and Ebadzadeh, 2011). For example, Diaf has shown that in the case of human activity recognition based on the combined KTH-Weizman dataset, Sigmoid and Log kernel functions have demonstrated superiority over other kernels including the Radical Basis Functions represented by Gaussian, Exponential, and Laplacian, (Diaf et al., 2012).

Nonparametric Discriminant Analysis provides a unified view of the Parametric Nearest Mean Reclassification Algorithm and the Nonparametric Valley Seeking Algorithm. Diaf combined NDA and KDA to introduce a non-parametric Fisher's discriminant analysis with kernels, (Diaf et al., 2013). Weighted LDA is commonly used in handling the unbalanced sample, (Jarchi and Boostani, 2006). Nearest Neighbor Discriminant Analysis (NNDA) can be regarded as an extension of NDA using a new between-class scatter matrix (Qiu and Wu, 2006). Above discriminant analyses are Parametric and Nonparametric Methods with a Global Classifier, which more or less identify nonlinear features.

In 2011, Fan proposed a parametric discriminant analysis with a local classifier in 2011 named Local Linear Discriminant Analysis (LLDA), which is skilled in complex nonlinear structures, (Fan et al., 2011). For each testing sample, LLDA first extracts the k-nearest subsets from the entire training set and then classifies them by LDA. The k-nearest subsets are calculated by Euclidean distance. (Shi and Hu, 2012) presents LLDA utilizing a composite kernel which is derived from a combination of local linear models with interpolation. Li et al proposed NDA with kernels, and tested the feasibility of the proposed algorithm on 3D model classification, (Li et al., 2013). (Zeng, 2014) proposed weighted marginal NDA to efficiently utilize the marginal information of sample distribution (Zeng, 2014).

NDA has been extended to a semi-supervised dimensionality reduction technique to take advantage of both the discriminating power provided by the NDA method and the locality-preserving power provided by the manifold learning (Du et al., 2013). (Du et al., 2013) embedded sparse representation in NDA for face recognition. Adaptive slow feature discriminant analysis is an attractive biologically inspired learning method to extract discriminant features for classification on time series (Gu et al., 2015). Fast incremental LDA feature extraction are derived by optimizing the step size in each iteration using steepest descent and conjugate direction methods (Ghassabeh et al., 2015).

(Li et al., 2017) generalizes LLDA to Local Kernel Nonparametric Discriminant Analysis (LKNDA), which is a Nonparametric Discriminant Analysis with a Local Classifier in which he showed that LKNDA performed more accurately and robustly than LLDA in almost all cases. LKNDA improves conventional Discriminant Analysis with the inspiration from Nonparametric Statistics. Their analysis considered the weight of different samples in subsets and modifies the Kernel function of Nonparametric statistics into a unilateral Kernel function. The Classification Function that they proposed relaxed the Normality Assumption, and performed well in the Nonlinear or Nonparametric Problem. Compared with the KNN method, they also showed that LKNDA has the same time complexity and higher accuracy on class margin.

From these reviews on Parametric and Nonparametric Discrimination Function, it may be seen that Parametric Discriminating Functions have shown good performance in situations where Linearity and Normality assumptions have been met. In cases where these assumptions are defied, Nonparametric Methods generally Perform better. It is also emerging from these reviews that even in cases where Parametric Methods are a good fit, Nonparametric ones may also deliver similar results except that their coverage rates are lower thereby depicting an aspect of robustness over the parametric ones.

Thus, in this study, a Nonparametric Kernel Discriminant Function is used in classifying the stateless communities in Kenya into the most appropriate local community(ies) using the 2009 Kenya Population and Housing Census data.

2.2.2 Classification Rates as a Measure of Robustness

In discriminant analysis each potential sampling unit has associated with it an index I of group membership, and a vector x of observation variables. A preliminary sample of units is collected and used to derive an estimator

$$\hat{I} = f(x)$$

of group membership. This estimator is based on the "discriminant" or "classification" functions, and it is subsequently used for classification of new samples. The pair of indices I and $\hat{I}$ result, and when they correspond for a given sampling unit it is said to be correctly classified. It is of interest to assess the frequency of this event, that is, to determine the correct classification rate.

Since agreement between I and $\hat{I}$ can occur by chance alone, irrespective of the discriminant functions, it is also important to adjust for chance agreements between I and $\hat{I}$.

Consider the usual linear discriminant function (see, for example, Anderson, (1958), then

$$v = [x - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)]' S^{-1}(\bar{x}_1 - \bar{x}_2)$$

where $\bar{x}_i$ denote the vector mean of the ith sample, $i = 1, 2$ and S is the Sample covariance matrix. The observation x is classified into population one or two according to the size of the expression v. When it is equally likely that observations have come from each of the two populations, and when the cost of the two possible incorrect decisions may be taken to be equal, then an appropriate classification rule is that the observation is classified into population one or two according to whether v is positive or negative.

If the Population Parameters (mean and variance) are known, then the linear discriminant function which yields the highest proportion correctly classified is

$$u = [x - \frac{1}{2}(\mu_1 + \mu_2)]' \Sigma^{-1}(\mu_1 + \mu_2)$$

If an observation x is classified into the first or second population according to whether u is positive or negative, then the probability of correct classification for an observation from the first population equals the probability of correct classification for an observation from the second population.

When observations are equally likely to have arisen from either population, the overall probability of correct classification is maximized using u with positive values of u assigned to the first Population and negative values of u assigned to the second Population. This maximum probability of correct classification, which will be denoted as p_1 is

$$p_1 = \Phi\left(\frac{\nabla}{2}\right)$$

where ϕ denotes the c.d.f. of the standard univariate normal distribution, and where ∇^2, the Mahalanobis distance between the two populations, is defined by

$$\nabla^2 = (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2)$$

The quantity p_1 is the highest attainable probability of correct classification. It is a parameter of the problem, and as such usually unknown.

Therefore, the need to classify individuals into one of two or more observed groups based upon asset of predictor variables is very common in the social and behavioral sciences (Zigler and Phillips, 1961). Due to the widespread use of classification methods, as well as the potential for errors in the initial classification of members in the training sample, and the important decisions and consequences often associated with the group into which an individual might be placed using these methods (Sireci et al., 1999); (DiStefano and Morgan, 2011), it becomes of utmost importance to determine not only which statistical classification methods are most accurate for the situation at hand but also which are most accurate when initial "true" group classifications may be questionable.

Previous researches indicate classification accuracy generally increases with increased sample size (Holden and Kelley, 2010), discrepancy in group size (Lei and Koehly, 2003); (Holden and Kelley, 2010), group separation (Blashfield, 1976); Lei and Koehly (2003); (Holden and Kelley, 2010), and number of variables used in the classification (Breckenridge, 2000). Assumption violations (Lei and Koehly, 2003), outliers and presence of Multi-Collinearity, (Pai et al. 2012), generally lead to decreased classification accuracy. Statistical classification can also be complicated by initial observed group misclassification (McLachlan, 1972); (Chhikara and McKeon, 1984); (Höfler, 2005); (Holden and Kelley, 2010) and (Balamurali and Kalyanasundaram, 2011).

Misclassification can be thought of as a type of measurement error (Ozasa, 2008) and can take several different forms. For example a distinction can be made between classification that occurs completely at random and misclassification that is non-random, occurring systematically based on the relative location of the point on its distribution (Lachenbruch, 1966), (Lachenbruch, 1974), (Chhikara and McKeon, 1984); (Holden and Kelley, 2010). Misclassification can also be differential or non-differential.

Non-differential misclassification occurs when the probability of misclassification is the same for all the study groups. Differential misclassification occurs when the probability of misclassification differs between study groups (Ozasa, 2008). Misclassification can also happen at either the exposure (for example, was the individual in the treatment or the control group) or the outcome level (Höfler, 2005); (Ozasa, 2008).

Given that each classification method has its strengths and limitations and that real world problems do not always satisfy the assumptions of a particular method, one approach when comparing different classification methods is to apply all appropriate methods and select the one that provides the best solution. This approach works well if, for a given problem situation, there is always one method (i.e., method A) that dominates all the others. Recent studies in comparing the performance of different classification techniques have been based mainly on experimental approaches (Almuallim and Dietterich, 1994) and (Dietterich et al., 1995), (Dietterich, 1995). Empirical comparisons among different algorithms suggest that no single method is best for all learning tasks (Salzberg 1991). In other words, each method is best for some, but not for all tasks.

There are different performance measures that are considered in literature to compare different classification functions. These measures include Accuracy which indicates the percentage of correctly classified instances during the course of classification, Misclassification rate which is the percentage of incorrectly classified instances are nothing, but the misclassification rate of the classifier and the Root mean squared error (RMSE) which usually provides how far the model is from giving the right answer. It represents the average prediction error within the same scale or unit (Panigrahi and Borah, 2018). A classification technique with the highest accuracy and with the lowest misclassification rate and root mean squared error is considered to be the most intelligent classifier for classification purposes.

(Lei and Koehly, 2003), Compared the relative accuracy of two widely used classification procedures, linear discriminant analysis and logistic regression, under various commonly encountered and interacting conditions. They investigated the effects of degree of group separation, equality of covariance structures, sample size, population prior, cut-score, and method of classification on total as well as separate-group classification errors using simulated multivariate normal data. Monte-Carlo Simulation was used to manipulate four factors under multivariate normality and the mean misclassification error rates and associated standard deviations computed and compared for various datasets. Their results showed that Logistic Regression(LR) and Linear Discriminant Analysis performed similarly. (Aye, 2021) employed discriminant analysis to predict undergraduate student's performance in the University system using data on the first year students of the department of Mathematical Science. The data was analysed using predictive discriminant analysis which yielded a canonical discriminant function that successfully predicted 87.2% of the graduating student's class of degree. The cross validated classification showed that overall 72.27% were correctly classified. The model performance was using Misclassification rates and showed lower misclassification rates indicating its accuracy.

From these reviews it can be seen that Nonparametric Classification Functions have higher Classification Rates compared to their counterpart Parametric Classification Functions. As is usually the case, a function with highier classificcation rate is considered to be a better classifier. Accordingly, Classification Rates shall be used as a measure of Robustness in this thesis especially at the empirical study stage.

2.3 Review of Statelessness

When one does not exist by law, he/she is extremely vulnerable to abuse and exploitation. Vulnerable people can easily be manipulated and even trafficked. Such people would suffer a whole lot of human rights violations. To limit such occurrences, every person should be entitled to a nationality. States have the sovereign right and duty to determine Nationality. Nationality is a particularly sensitive issue for countries and has often led to intractable disputes, tension and conflicts. Many states have resisted international laws and obligations-including those they have agreed to. This, they argue is in the spirit of defending their sovereign rights. Specifically, governments cling to the position that decisions around nationality are a matter of national sovereignty and that it is up to the state to bestow or withhold nationality from any individual living within its territory.

However, in July 2019, Kyrgyzstan became the first country in the world to eradicate statelessness, primarily through documentation drives that targeted minority populations, (UNHCR, 2022). Up to 25,000 children born to Venezuelan parents who have fled to Colombia amid a political and economic crisis in their homeland may be stateless or are at risk. Children must have at least one Colombian parent to qualify for citizenship. In a measure to combat undocumented refugee babies, Colombia granted citizenship to more than 24,000 children born to Venezuelan migrants on its territory since 2015, and to those who were born there before 2021.

Statelessness poses serious threats to development, public health, security and international relations, (Tucker, 2021). It is not only a source of human insecurity and a cause of forced displacement but also poses real threats to national and regional stability. Statelessness may occur for a variety of reasons including discrimination against particular ethnic or religious groups or on the basis of gender; the emergence of new States and transfers between existing States; and conflict of nationality laws. Statelessness is often the product of policies that aim to exclude people deemed to be outsiders, notwithstanding their deep ties to a particular country, (Manby, 2011).

The exact number of stateless people globally is not known, but United Nations High Commission for Refugees (UNHCR) estimates the number to be about 10 million. At the international level, more than 900,000 people in Myanmar's Rakhine state are stateless on the basis of the current citizenship law, which provides that only members of certain ethnic groups are eligible for citizenship. Specifically, in 1982, Buddhist-majority Myanmar passed a citizenship law that effectively rendered Rohingya Stateless, (UNHCR, 2017). They are Muslims and of South Asian descent. Ethnic violence has driven many to leave, but hundreds of thousands still remain in Myanmar. There are about 900,000 Rohingya in neighboring Bangladesh and smaller populations across Asia.

In addition, some 25 States around the world do not allow women to transfer nationality to their children, statelessness can occur where fathers are unknown, missing or deceased, (UNHCR, 2017). Areas that have experienced large-scale displacement have also been significantly affected by statelessness. Statelessness due to the dissolution of former states also continues to affect many people, including some 600,000 people in Europe alone, (UNHCR, 2022). There have been notable examples where, through political will, it has been possible to resolve large protracted situations of Statelessness. For example, the case of some Urdu-speakers was resolved in Bangladesh in 2008, (UNHCR, 2022). Similarly, the situation of the Brasileirinhos Apatridás, stateless children born to Brazilian parents abroad who were unable to acquire Brazilian Nationality unless they went back to live in Brazil, was resolved in 2007, (UNHCR, 2022).

Estonia has recently took steps to further facilitate the acquisition of Citizenship by those born in each of the two countries to non-citizen parents, to ensure that these situations are resolved. It was done by amending its Citizenship law to make it easier for several category of people including children to become naturalized by January 2016, (Evas and Väljataga, 2016). Nearly 479,000 people are Stateless in Thailand, including members of ethnic hill tribes such as the Yao, Hmong and Karen who live in the mountainous border with Myanmar and Laos and with the semi-nomadic 'Sea Gypsies' along the Andaman coast. When the Soviet Union broke up, many ethnic Russians were stranded in the new Baltic states (Estonia and Latvia) and defined as "non-citizens". As a result, about 225,000 stateless people live in Latvia and 78,000 in Estonia, mainly ethnic Russians who have trouble obtaining citizenship and often face discrimination.

In 1962, many Kurds in the North-East were stripped of Citizenship in Syria, a move considered to "Arabise" the resource-rich region. Before the civil war, there were an estimated 300,000 stateless Kurds in Syria, (Lynch and Ali, 2006). However, the UN data suggests the number of stateless fell to 160,000, (Albarazi, 2016). Experts on Statelessness have warned that babies born to Syrian refugee women in Lebanon and Jordan could end up being stateless. In Kuwait, many people among the nomadic Bedouin tribe failed to acquire Citizenship at independence in 1961. Their descendants are known as Bidoon, which means "without" nationality in Arabic. There are about 92,000 Bidoon in Kuwait, according to U.N. data who are still stateless. They are often barred from free education, healthcare and many jobs, (Van Wass, 2010).

In Iraq there are about 47,500 Stateless people who include Bidoon, Palestinian refugees and Faili Kurds, an ethnic group that historically live on both sides of the Iraq-Iran border. At least 150,000 Faili Kurds had their Nationality revoked in 1980 under the Bath regime. Although many have since had their Nationality reinstated, some remain stateless, (Information et al., 2022). Nepal argues that it does not have a stateless population but experts on Statelessness believe many people, may be affected. Part of the problem is derived from a law banning women married to foreigners passing on Nationality to their children. There is also a Stateless population of people who were expelled by Bhutan in the 1990s.

In the year 2013, Dominican Republic made a court ruling that was aimed at tackling illegal migration and this left many people stateless. These are people mostly of Haitian descent who were born in Dominican Republic, (AmnestyInternational, 2015). In 2015, they were about 134,000 stateless people, according to U.N. data.

There are tens of thousands of stateless Roma, an ethnic group with origins in India who live in central and eastern Europe. With the break-up of Czechoslovakia and Yugoslavia, successor states claimed they belonged elsewhere. Other Roma in Kosovo and Bosnia have become stateless due to war-time displacement. Roma families often do not register their children's births or hold official property titles, preferring to pass houses to relatives informally. This makes it hard to prove where they are from.

Statelessness is increasingly being recognized as a major problem in Africa. However, it not properly documented because the official stateless population significantly overlaps with a much larger Population of undocumented people who are unaware of their official nationality status. The very nature of Statelessness, that is, people are undocumented and unaccounted for makes it hard to know exactly how many people in Africa are affected. In most African states, nationality laws are based on the concepts of jus soli, or 'right of soil,' and jus sanguinis, or 'right of blood.' Under the former, the person can obtain citizenship if they are born in that particular country, while the latter bases a person's nationality on the origin of their parents.

In many cases, States which primarily base their Nationality laws on the principle of jus soli prevent populations who are away from their 'historic' homeland to apply for citizenship of that country, while at the same time being denied nationality of their country of residence due to laws based on jus sanguinis. They are in limbo, because they are not protected by the Citizenship of their new country and at the same time they are not protected by their country of origin because they are no longer considered as Citizens.

According to UNHCR, Cote D'Ivoire, Zimbabwe and Kenya have an estimated 700,000, 300,000 and about 20,000 stateless persons respectively. At the end of 2015, UNHCR recorded 1,021,418 persons under its statelessness mandate in Africa, but the real figure is probably much higher as this is based on the estimated populations in only six countries, (AmnestyInternational, 2017).

Statelessness in Africa has a number of causes, (AU, 2018). Of the 27 States in the world which still discriminate against women in their ability to transmit nationality to their children, 9 are in Sub-Saharan Africa. A worrying trend is that many African States do not have safeguards guaranteeing nationality to children born in their territory who would otherwise be stateless, with the result that children continue to be born stateless across Africa. Racial, Religious, and Ethnic Discrimination are present in the Nationality Laws of around ten African States and result in individuals being unable to acquire nationality. Nomadic and cross-border populations continue to face practical and political challenges as Nationality Laws are not designed to accommodate them and settled populations remain suspicious of their loyalties.

Displaced persons, including refugees, run the risk of losing their connection with their country of origin as well as facing difficulties acquiring documentation, which may result in Statelessness, particularly in subsequent generations. State succession, both the legacy of decolonization and more recent succession situations, and the resulting redefinition of National belonging are also a cause of Statelessness in Africa. Statelessness can also result from lack of due process and the broad discretion granted to State officials responsible for the issuing of birth certificates and identity cards, which in practice may determine an individual's access to nationality.

Among the countries in Africa where UNHCR recognizes that there are major populations at risk of Statelessness are Côte d'Ivoire, the Democratic Republic of Congo, Eritrea, Ethiopia, Kenya, Madagascar, South Africa, Sudan and Zimbabwe. In Côte d'Ivoire and DRC the failure to recognize the nationality of large populations belonging to particular Ethnic Groups has been one of the main causes of conflict in those countries.

An estimated 750,000 Stateless Persons live in 15 countries that make up the West African region. The region is also home to a large population of persons at risk of statelessness. Cote d'Ivoire has a large stateless populations, many of whom were migrants of Burkina Faso, Mali and Ghana, who were encouraged to work on coffee and cotton plantations in Ivory Coast in the 20th century. They were not eligible for Ivorian Nationality after the Country's Independence from France in 1960.

Four of the nine African countries with the biggest stateless populations are in Southern Africa: Zimbabwe, South Africa, Madagascar and the Democratic Republic of the Congo. Statelessness in Southern Africa is driven primarily by colonial history, border changes, migration, poor civil registry systems, and discrimination on the basis of gender, ethnicity and religion. Zimbabwe has been called the 'main' statelessness crisis in Southern Africa.

The Statelessness Situation in Zimbabwe has persisted for generations as a means of political exclusion, with little signs of improvement. Many farm workers are of foreign African origin, although most are born in Zimbabwe. For decades, the government has implemented a series of complex citizenship rules to prevent these farm workers from voting. This has led to statelessness and hardship for some of the country's most vulnerable people. In South Africa, the situation threatens to worsen due to rising nationalism and anti-migrant sentiments and it appears to be on a path to continue using nationality as a weapon, further deepening xenophobic sentiments towards other Africans. Most of the stateless population in South Africa are migrants, asylum seekers and refugees from the region.

Kenya is a home to different groups of stateless persons such as the, Pemba, Galjeel as well as groups of individuals of Burundian, Congolese, Indian and Rwandan descent. People of Kenyan Somalis whose access for Kenyan identification documents has been limited are likely to face the same situation of statelessness.

The number of stateless persons in Kenya is not clearly known. However, it is estimated to be 18,500 after the registration of the Makonde in 2016.

In 2016, UNHCR in cooperation with Statistics Norway, Kenya National Bureau of Statistics (KNBS) and Haki Centre conducted a survey of the Pemba people living in Kwale and Kilifi counties. In 2019, UNHCR in coordination with the Government of Kenya through KNBS conducted a survey of the Shona in Kenya. UNHCR continues to raise awareness in Kenya on the issue of statelessness through media, community forums as well as sensitization of relevant stakeholders with the aim of resolving existing statelessness situations. This includes improving the access to Kenyan documentation (birth certificates and national ID cards) especially in Kwale, Kilifi, Garissa and Kiambu counties where majority of stateless persons live in. This helps to avoid having children who are undocumented and may become stateless later in life. The move is also to boost the achievement of the relevant Sustainable Development Goals Target 16.9 aim to provide legal identity for all including free birth registrations by 2030.

Galjeel Somalis is a group that has lived in Kenya for decades but are not regarded as Kenyans. In 1989, a government action separated Kenyan Somalis from those regarded to be from Somalia and so, many Galjeel Somalis were branded non-Kenyans, and their identity cards were confiscated. The Galjeel were forced to move to remote areas of the country including Tana River, with very little or no basic amenities hence are deprived of basic human rights.

According to UNHCR, there are three most applicable solutions for persons of concern (refugees and stateless persons): voluntary repatriation, local integration and resettlement. Voluntary repatriation is usually the best option for persons of concern as long as it is done voluntarily, in safety and with dignity. This option however is out of question for stateless persons which leaves local integration as the second best option for the persons of concern.

According to Carciotto and Christiano, (2017), local integration process involves three interrelated dimensions: First dimension involves a greater range of socio-economic rights and entitlement to the host-state. The second dimension is that local integration can be regarded as an economic process which involves livelihoods so as to attain a growing degree of self-reliance. A third dimension is that local integration initiates a social process which in essence enables persons of concern to live in a pacific environment together with the host population, without fear of systematic discrimination, intimidation or exploitation by the authorities of the host country. The other option is third country resettlement that involves tripartite agreements between governments and other bodies such as relevant UN agencies.

Having noted that local integration is an economic process which involves livelihoods so as to attain a growing degree of self-reliance, economic and social rights of stateless persons is mostly at risk. Economic and social rights are put in place in order that people can live, work and develop to their fullest potential as human beings, (Coomans, 2005). The Governments world over have strived to provide a conducive environment so that their citizens achieve economic development.

Article 11 of the International Covenant of the Economic and Social Rights alludes to the fact that State parties shall recognize the right of everyone to an adequate standard of living for the individual and his family, which encompasses other rights such as right to adequate food, clothing and housing and to the continuous improvement of living conditions (UN, 1966). For these rights to be operationalized, there should be no people categorized as being stateless.

2.4 Summary of the Literature Review

From the literature reviewed in this chapter, we take note that KNN is a competent nonparametric classification algorithm that is capable of solving the problem of complex nonlinear characteristics with adequate sample size. However, KNN is not capable of understanding the law of the points near the class interface. LDA is a linear projection classifier and it performs well with a small sample size but is invalid in a nonlinear environment. SVM can solve simple nonlinear problems by establishing a classification hyper plane but it is still insufficient to handle complex nonlinear problems. LKNDA strives to absorb the advantages of both nonparametric and parametric methods. It takes the advantages of nonparametric methods in solving complex nonlinear problems. At the same time, it draws on the benefits of parametric ones for pattern recognition with a small sample size. If KNN behaves poorly, or if you want to further improve the prediction accuracy, LKNDA becomes a good choice. However, this study is to focus on kernel based nonparametric approach.

There are two different definitions of kernel function: in the field of machine learning, a kernel function is used to map a sample set into a high dimensional space; in the field of nonparametric statistics, a kernel function is a weighting function of nonparametric estimation. In this study, we exploit the strength of the first definition in the nonparametric approach and apply it in classifying stateless persons in Kenya.

2.5 Research Gaps Resulting from Critiques of the Existing Literature Reviewed

From the literature reviewed in Sections **??**, 2.2.2 and 2.3 where the critiques have been given, it was evident that Parametric Discriminant Functions such as Linear and Quadratic Discriminant Functions have extensively been used in Classification problems. The two methods have been relied upon by many researchers. However, these conventional parametric classification methods face difficulty in addressing the Non-Gaussian aspects of Sample distributions due to their Parametric nature of scatter matrices. It is for these reasons that we shift our focus to the Nonparametric approach because it is able to overcome the assumptions of linearity and normality. Whereas researches have been done in the area of Nonparametric approach, there is no known work that has exploited the application of Nonparametric kernel discriminant function to classify objects. This study proposes a Nonparametric Kernel Discriminant Function to solve the problem of stateless persons in Kenya by applying various relevant datasets available.

CHAPTER THREE
METHODOLOGY

3.1 Introduction

Classification of information is an important component of all decision-making tasks. In such tasks, there are instances which can easily be formulated into classification problems. One of such problems in our world today is the classification of stateless people in an amicable way so that problems associated with statelesness are addressed. This classification calls for a Robust Function whose Discrimination power does not suffer from problems faced by Parametric Discriminant Functions.

To enable development of a Robust Discriminant Function, a brief review of the traditional Discriminant Functions is now done in Section 3.2 that follows. Thereafter, reference is made to section 3.3 where a review of Kernel Functions was done with an intention of highlighting reasons towards choice of the Epanechnikov Kernel that is eventually used in this development. From Table **??**, among other known theoretical reasons, this Kernel function is used in this work since it has the highest efficiency. Finally, some notations used are provided and briefly explained.

3.2 Discriminant Functions and Classification

3.2.1 Bayes Discriminant Rule

Suppose we have a set of v populations or groups that correspond to density functions $f_1, f_2, \ldots, f_v$. Our aim is to assign all points x from the sample space to one of these groups or densities. We compare the weighted heights of the density functions to obtain the Bayes Discriminant Rule

$$x \ is \ allocated \ to \ group \ j_0 \ if \ j_0 = arg \max_{j \in 1,\dots,v} \pi_j f_j(x) \tag{3.1}$$

where π_j is the prior probability of drawing from density f_j. If we enumerate for all x from the sample space, we produce a partition $P = \{P1, P2, \dots, P_v\}$ of the sample space using

$$x \in P_j \ if \ x \ is \ allocated \ to \ group \ j$$

The Discriminant Rule, Equation (3.1), contains the unknown density functions and the (possibly) unknown prior probabilities. Once we collect some data, we can modify this abstract rule into a practical one.
rr We collect training data $X_j = \{X_{j1}, X_{j2}, \dots, X_{jn_j}\}$, drawn from f_j , for $j = 1, 2, \dots, v$. (The sample sizes n_j are known and non-random).

A priori there is a class structure in the population since we know which data points are drawn from which density function. From these training data, we can construct a practical discriminant rule and subsequent partition.

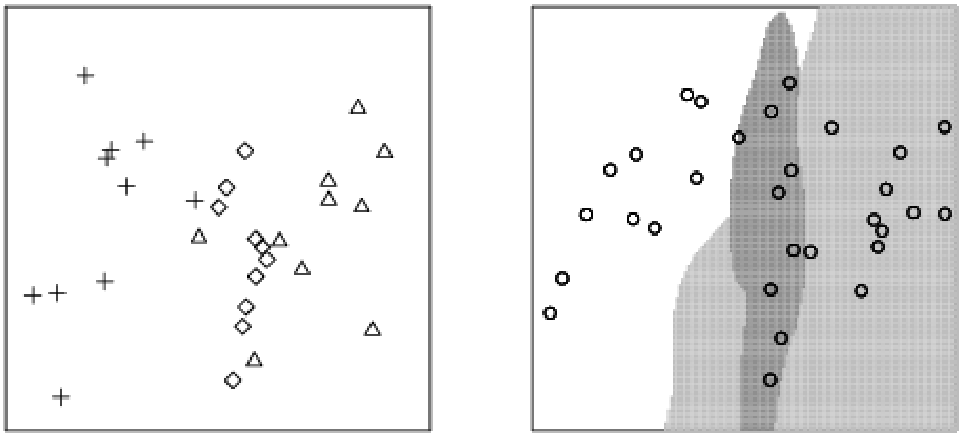
Using this discriminant rule/partition, we classify the test data $Y_1, Y_2, \dots, Y_m$, drawn from

$$f = T = \sum_{i \in s} y_i + \sum_{j=1}^{v} \pi_j f_j(x)$$

This time, we do not know which populations generated which data points.

An illustration of partitioning and discriminating using this Bayes discriminant rule into three groups is given in Figure 3.1 . There are three training sets, each of size 10, denoted by the pluses, diamonds and triangles on the left diagram. The prior probabilities are equal to 1/3.

The three (normal) density functions (not shown) are compared according to Equation (3.1) and this yields the partition on the right: white-pluses, dark grey-diamonds and light grey-triangles. The circles are the 30 test data points that we are attempting to classify.



(a) (b)

Figure 3.1: Partition and Discrimination from Discriminant Analysis: Plus-White, Circle-Dark Grey, Triangle-Light Grey, Circles are Test 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. In addition, the fundamental assumptions made by the traditional Bayes Discriminant Rule is that all the features are independent of one another and contribute equally to the outcome i.e all are of equal importance. But these assumptions are not always valid in real life. Also the Bayes Rule does inform on how to select the prior probabilities hence creating the prior distribution is difficult. In Nonparametric Discriminant Analysis, we relax this assumption and thus are able to tackle more complex cases. We will focus on Kernel Methods for discriminant analysis. The Monographs of (Silverman, 2018), (Scott, 1991) and (Simonoff, 1996)(Chapter 7) contain summaries of Kernel Discriminant analysis while Hand (1982) contains more detailed and lengthy expositions on this subject.

3.2.2 Parametric Discriminant Analysis

The two Parametric Methods that we describe in more detail here, linear and Quadratic Discriminant analysis, are among the most commonly used. Their ease of computation is as a result of some underlying Normality assumption.

Linear Discriminant Functions

We assume that the densities f_j are normal with different mean vectors μ_j and with common variance matrix Σ. The key assumption is that $f_j \sim N(\mu_j, \Sigma)$. After taking the logarithm of f_j, the discriminant rule, equation (3.1), reduces to

$$\begin{gathered} x \text{ } is \text{ } allocated \text{ } to \text{ } group \text{ } j_0 \text{ } if \\ j_0 = arg \max_{\{j \in 1,\ldots,v\}} \log(\pi_j) - \frac{1}{2}(x-\mu_j)^T \Sigma^{-1}(x-\mu_j) \end{gathered} \tag{3.2}$$

From this equation, we can see that resulting partition is obtained by intersections of ellipsoids with different centres and with the same orientation. This yields partition boundaries that are hyperplanes. For our example data from Figure 3.1, we apply the linear discriminant rule to obtain the partition in Figure 3.2, using the sample mean $\bar{X}_j$ as estimate of μ_j and $S = (n-v)^{-1} \sum_{i \in s} y_i + \sum_{j=i}^{v} n_j S_j$ for Σ where S_j is the sample variance.

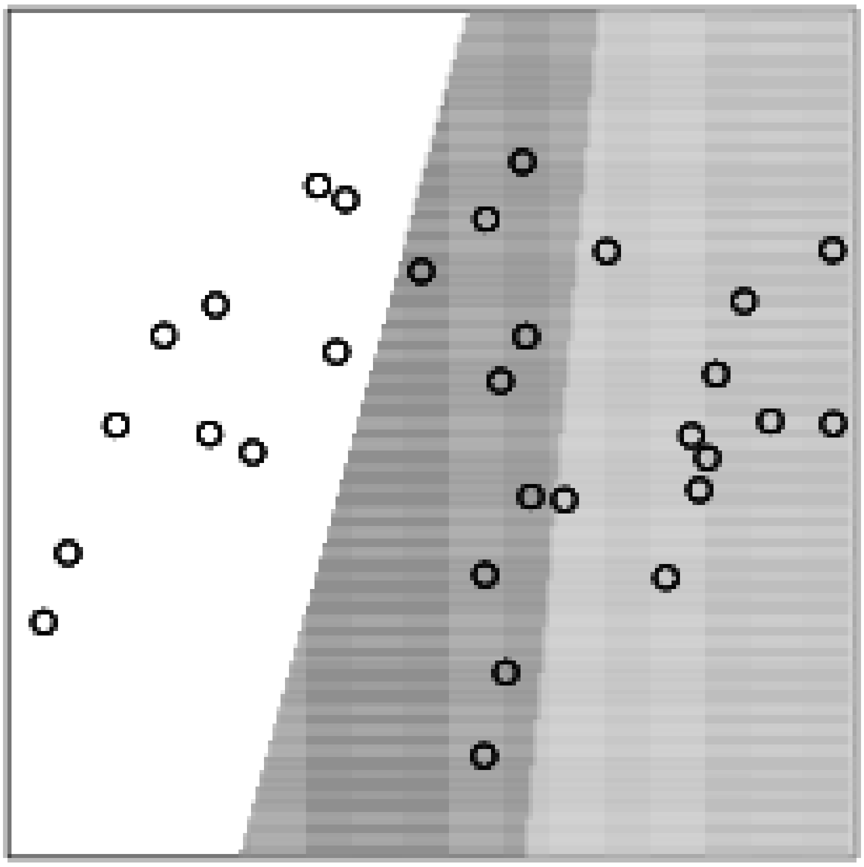


Figure 3.2: Partition from Linear Discriminant Analysis

Quadratic Discriminant Functions

For Quadratic Discriminants, we have that the densities are normal with different means μ_j and different variances Σ_j . The assumption of common variance of linear discriminant analysis is relaxed. That is , $f_j \sim N(\mu_j, \Sigma)$. The discriminant rule, Equation (3.1), reduces to (after taking logarithms of f_j)

$$\begin{aligned} &x \text{ } is \text{ } allocated \text{ } to \text{ } group \text{ } j_0 \text{ } if \\ &j_0 = arg \underset{\{j \in 1,\ldots,v\}}{max} \log(\pi_j) - \frac{1}{2}\log|\Sigma_j| - \frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j) \end{aligned} \tag{3.3}$$

This discriminant rule yields a partition defined by intersections of ellipsoids with differing centers and orientations. The boundaries are thus piecewise paraboloidal curves, as is illustrated in Figure 3.3, obtained by replacing the means and variances with their sample statistics.

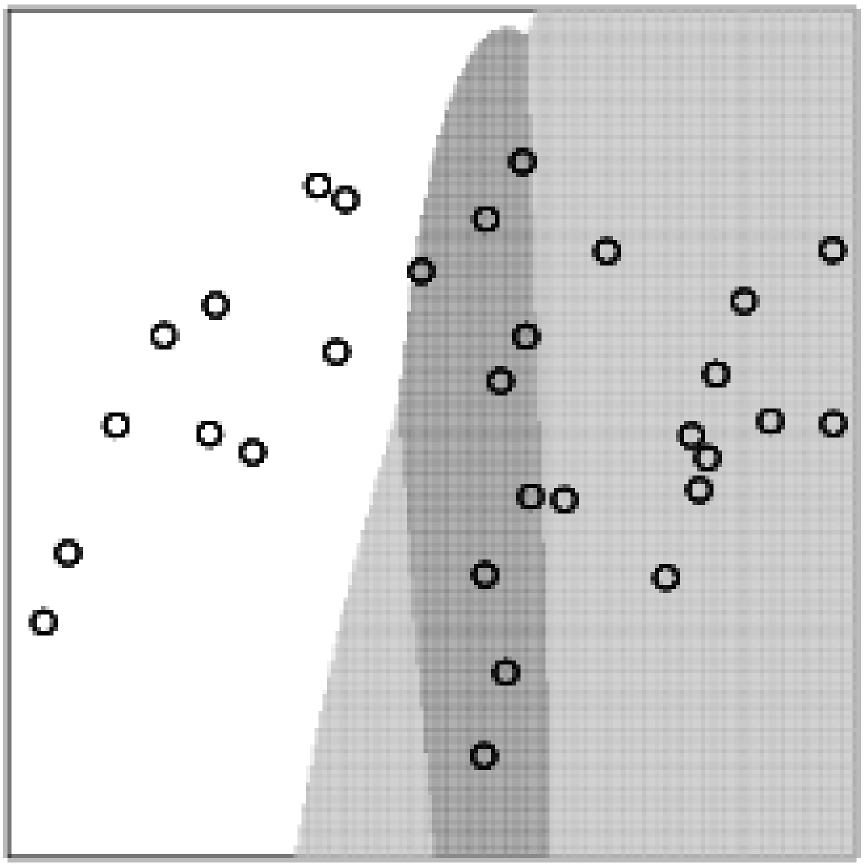


Figure 3.3: Partition from Quadratic Discriminant Analysis

To effectively use the parametric discriminant rules, one has to replace the unknown parameters with their usual sample estimates.

3.2.3 Robust Nonparametric Kernel Discriminant Function

In this section, the Robust Nonparametric Kernel Discriminant Function that is used in this thesis is developed. It is noted that Parametric Methods can be generalized to Nonparametric ones. In this development, the traditional Bayes Discriminant Function was adopted as a basis for Discrimination except that the parametric Functions are replaced with the appropriate Kernels.

In the development of our Discriminant Function, we start from equation 3.1 which is the general Bayes Discriminant Rule, at the point of density function is used, a Kernel Density Function used instead of Parametric Density Function. Thus, let J be a group of the Population in which an individual is to be classified into, f_j the Density Function, $K()$ be the Kernel Function. Also we let, S to be the training sample and n its Sample Size. Let H_j be the bandwidth and π_j the prior probabilities.

With the foregoing notations, Kernel density estimation, (Scott, 1992), (Silverman, 2018) is a popular method for Nonparametric Density Estimation, and it has one well known application in Kernel Discriminant Analysis (KDA), (Hall and Wand, 1988). In a J class classification problem, if there is a Training Sample $S = \{(x_i, c_i)\,; x_i \in \mathbb{R}^d, C_i \in (1, 2, \ldots, J), i = 1, 2, \ldots, n\}$ of n observations, the Kernel Estimate for the Density Function $f_j(j = 1, 2, \ldots, J)$ can be expressed as

$$\hat{f}_{jh}(x) = \frac{1}{nh^d} \sum_{i:c_i=j} K\left\{\frac{1}{h}(x - x_i)\right\} \tag{3.4}$$

where n_j is the number of observations from the jth class $\sum n_j = n$ and 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 used to construct the Kernel Discriminant Rule (KDR) given by

$$KDR:\ x\ is\ allocated\ to\ group\ j_0\ if\ j_0 = arg \max_{j\in\{1,\ldots,v\}} \hat{\pi}_j \hat{f}_j(x, H_j) \tag{3.5}$$

where $\hat{f}_j(x, H_j)$ is the Kernel Density Estimate corresponding to the jth group and where π_j is the prior probability of the jth group. If these priors are not known, one usually estimates them using training sample proportions $\hat{\pi}_j = \frac{n_j}{n}$, $(j = 1, 2, \ldots, J)$ of different groups. Many choices for the kernel function K are available in the literature, (Scott, 1992), (Silverman, 2018). Equation 3.5 forms our proposed classification rule. See Figure 3.4 for Illustration of plug-in bandwidth selectors for H_j.

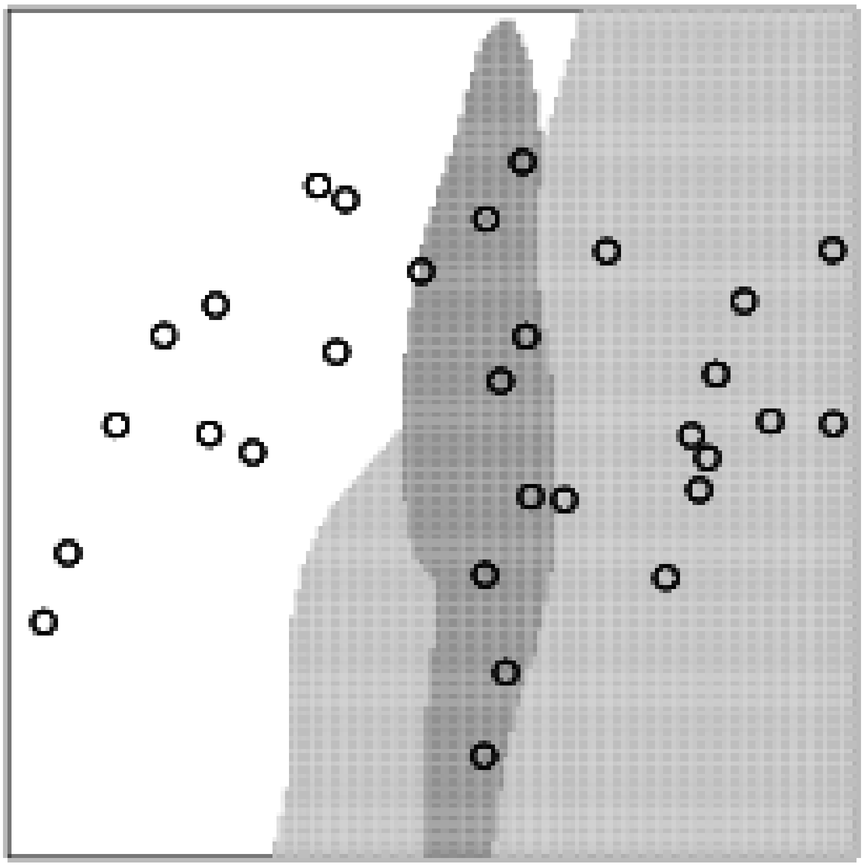


Figure 3.4: Partition From Kernel Discriminant Analysis

Now that we are using Kernel Density Estimators for discriminant analysis, selection of appropriate bandwidths is crucial. (Hand, 1982) contains discussion on this question. On one hand, we can attempt to find Optimal bandwidths for Optimal individual kernel density Estimates and on the other hand, we could find Optimal Bandwidths which directly optimize the Misclassification Rate (MR), as (Hall and Wand, 1988) attempt for the two.

3.3 Nonparametric Kernel Density Functions and Bandwidth Selection

Kernel Density Functions

A Kernel is a Mathematical F unction that returns a probability of a random variable for a given value. It is any smooth function k such that $k(u) \geq 0$ and $\int k(u)du = 1$, $\int uk(u)du = 0$ and $\sigma_u^2 = \int u^2 k(u)du > 0$. A kernel function weights the contribution of observations from a data sample based on their distance to a given Sample. A parameter called the "bandwidth" or "smoothing" controls the scope of observations from the data sample that contributes to estimating the probability of a given Sample.

In practice, there are a number of kernel functions to choose from,however the most three commonly used kernels are:
Gaussian kernel

$$k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), -\infty < u < \infty \tag{3.6}$$

Epanechnikov kernel

$$k(u) = \frac{3}{4}\left(1 - u^2\right) I(|u| \leq 1) \tag{3.7}$$

Biweight or Quartic kernel

$$k(u) = \frac{15}{16}\left(1 - u^2\right)^2 I(|u| \leq 1) \tag{3.8}$$

Three other kernels that aren't as common are:
Uniform kernel

$$k(u) = \frac{1}{2} I(|u| \leq 1) \tag{3.9}$$

Triangular kernel

$$k(u) = (1 - |u|) I(|u| \leq 1) \tag{3.10}$$

Triweight kernel

$$k(u) = \frac{35}{32}\left(1-u^2\right)^3 I(|u| \leq 1) \tag{3.11}$$

Table 3.1: Various Kernels and their Efficiencies

Kernel	Function	Efficiency
Epanechnikov	$\frac{3}{4}\left(-u^2+1\right) I\left(\lvert u\rvert \leq 1\right)$	1.000
Biweight	$\frac{5}{6}\left(1-u^2\right)^2 I\left(\lvert u\rvert \leq 1\right)$	0.994
Triangular	$\left(1-\lvert u\rvert\right) I\left(\lvert u\rvert \leq 1\right)$	0.986
Normal	$(2\pi)^{-\frac{1}{2}} e^{-\frac{u^2}{2}}$	0.951
Uniform	$\frac{1}{2} I\left(\lvert u\rvert \leq 1\right)$	0.930
Triweight	$\frac{35}{32}(1-u^2)^3 I\left(\lvert u\rvert \leq 1\right)$	0.987

Source: (Jann, 2007)

In Table 3.1, the efficiency, of the kernel has been given relative to Epanechnikov kernel. Epanechnikov kernel is optimal since it minimizes asymptotic mean integrated squared error. From the information in Table 3.1, one can note that the choice of these kernels rarely affects the estimates in a significant way.

The choice of the kernel function determines the weight given to each observation. For instance, a Uniform Kernel Function assigns equal weights to all points closest to the target and diminishes the weights to those points that are "farthest" from the center of the kernel. According to (Pflug, 1996) performance of Kernel is measured by Mean Integrated Squared Error or Asymptotic Mean Integrated Squared Error.

Bandwidth Selection

The bandwidth of a kernel is a free parameter which exhibits a strong influence on the resulting estimate. Kernel smoothing requires the choice of a bandwidth parameter. This choice is critical, as under- or over-smoothing can substantially reduce precision.

To illustrate its effect, we take a simulated random sample from a random sample of 100 points from a standard normal distribution (plotted at the blue spikes in the rug plot on the horizontal axis) as shown in figure 3.5. The black solid curve is the true density (a normal density with mean 0 and variance 1). In comparison, the blue dotdash curve is under-smoothed since it contains too many spurious data artifacts arising from using a bandwidth h = 0.05, which is too small.

The green dotted curve is over-smoothed since using the bandwidth h = 2 obscures much of the underlying structure. The red dashed curve with a bandwidth of h = 0.5 and the grey two dashed curve with a bandwidth h=0.337 is considered to be optimally smoothed since its density estimate is close to the true density.

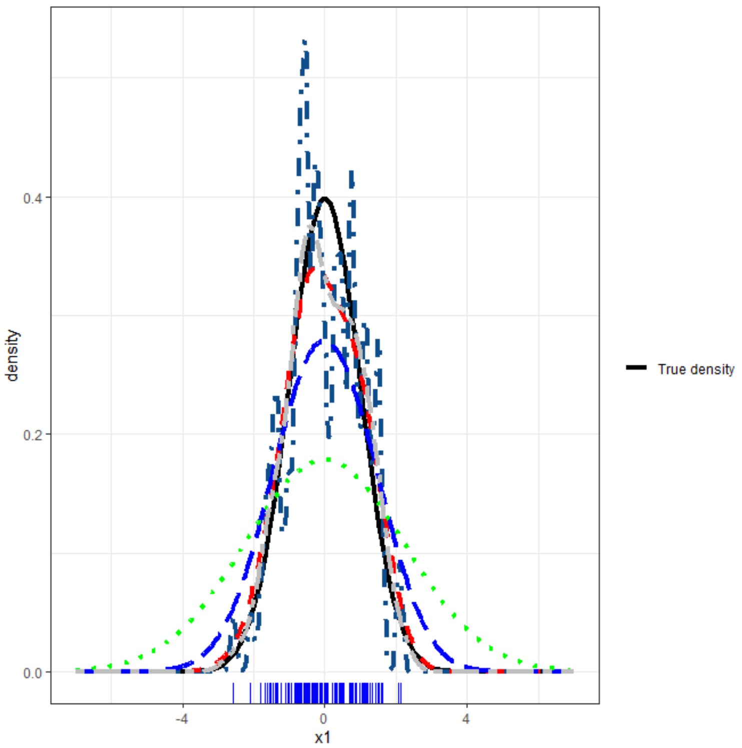


Figure 3.5: Kernel Density Estimate (KDE) with different bandwidths of a random sample

The most common optimality criterion used to select this parameter is the expected $L2$ risk function, also termed the mean integrated squared error:

$$MISE(h) = E\left[\int (f_h(x) - f(x))^2 \, dx\right]$$

Under weak assumptions on f and K, $MISE(h) = AMISE(h) + o(\frac{1}{nh} + h^4)$ where o is the little o notation. The AMISE is the Asymptotic MISE which consists of the two leading terms

$$AMISE(h) = \frac{R(K)}{nh} + \frac{1}{4} m_2(K)^2 h^4 R(f'')$$

Where $R(g) = \int g(x)^2$ a function g

$$m_2(K) = \int x^2 K(x) dx$$

To be able to prove the theoretical results, the following assumptions are made; and f'' is the second derivative of f. The minimum of this $AMISE$ is the solution to this differential equation

$$\frac{d}{dh} AMISE(h) = -\frac{R(K)}{nh^2} + m_2(K)^2 h^3 R(f'') = 0$$

or

$$h_{AMISE} = \frac{R(K)^{\frac{1}{5}}}{m_2(K)^{\frac{2}{5}} R(f'')^{\frac{1}{5}} n^{\frac{1}{5}}}$$

Neither the $AMISE$ nor the h_{AMISE} formulas can be used directly since they involve the unknown density function f or its second derivativef'', so a variety of automatic data-based methods have been developed for selecting the bandwidth. Many review studies have been carried out to compare their efficacies, h with the general consensus that the plug-in selectors and cross validation selectors are the most useful over a wide range of data sets.

Substituting any bandwidth h which has the same asymptotic order $n^{\frac{-1}{5}}$ as h_{AMISE} into the $AMISE$ gives $AMISE(h) = O(n^{\frac{-4}{5}}$, where O is the big O notation. It can be shown that, under weak assumptions, there cannot exist a non-parametric estimator that converges at a faster rate than the kernel estimator. Note that the $n^{\frac{-4}{5}}$ rate is slower than the typical $n - 1$ convergence rate of parametric methods.
If the bandwidth is not held fixed, but is varied depending upon the location of either the estimate (balloon estimator) or the samples (pointwise estimator), this produces a particularly powerful method termed adaptive or variable bandwidth kernel density estimation.

Further, the bandwidths vary with the kernel function chosen. An optimal bandwidth of one Kernel Function cannot be regarded in the same way for another Function. Because of this, many researchers have been carrying out studies aimed at determining techniques of obtaining bandwidths that minimize MSE or $AMSE$ functions that can be used with the different Kernel Functions.

Two methods, the "plug-in" and "cross validation" are the common ways in which this problem can be tackled. The plug-in method simply involves the replacement of the unknown functions in the expression of interest. The AMISE optimal bandwidth equation 3.12 depends on the unknown roughness R_1. A simple choice is a normal scale estimate. if $f = \phi_\sigma$

$$R_1 = \int_{-\infty}^{\infty} (\phi_\sigma^{(1)}(y))^2 = \frac{1}{\sigma^3} \int_{-\infty}^{\infty} y^2 \phi(y)^2 = \frac{1}{\sigma^3 4\sqrt{\pi}}$$

Thus, a reference bandwidth is

$$\hat{h}_0 = \hat{\sigma}(4\sqrt{\sigma\psi})^{\frac{1}{3}} n^{-\frac{1}{3}} \tag{3.12}$$

Where $\hat{\sigma}$ is the sample standard deviation. In particular, for the normal kernel $K = \phi$ then $\hat{h}_1 = 1.59\hat{\sigma}n^{-\frac{1}{3}}$. The reference bandwidth, however, may work poorly for distributions which are far from the normal. As shown by Jones (1990) if $hn^{\frac{1}{2}} \to \infty \ as \ n \to \infty$ then

$$AMISE(h) = \int_{-\infty}^{\infty} E\left(\hat{F}_h(y) - F(y)\right)^2 dy = \frac{V}{n} - \frac{h\psi}{n} + \frac{h^4 R_1}{4} + O(h^4) \tag{3.13}$$

where, $V = \displaystyle\int_{-\infty}^{\infty} F(y)\left(1 - F(y)\right)^2 dy$, $\psi = 2\displaystyle\int_{-\infty}^{\infty} xK(x)k(x)dx > 0$ is a constant which depends only on the kernel. For example, if $k(x) = \phi(x)$ then $\psi = \frac{1}{\sqrt{\pi}}$. The $AMISE$ is minimized by setting h equal to $h_0 = (\frac{\psi}{R_1})^{\frac{1}{3}} n^{-\frac{1}{3}}$. The optimal AMISE is

$$AMISE(h_0) = \frac{V}{n} - \frac{3\psi^{\frac{3}{8}}}{n^{\frac{3}{4}} R_1^{\frac{1}{8}}} \tag{3.14}$$

3.4 Misclassification Rate (MR)

After development of a Robust Nonparametric Discriminant Function, the study embarked on estimating its Classification Rates. This was done as a measure of its Robustness. To do this, the standard Bayes procedure was employed.

The Misclassification rate gives proportion of points that are assigned to an incorrect group based on a discriminant rule. Then we have

$$\begin{aligned}1 - MR =& P(Y \ is \ classified \ correctly) \\ =& E_Y \left[1\left\{Y \ is \ classified \ correctly\right\}\right] \\ =& E_X \left[E_Y \left[1\left\{Y \ is \ classified \ correctly\right\}\right] \ |X_1, X_2, \ldots, X_v\right] \\ =& \frac{TP + TN}{TP + FP + TN + FN}\end{aligned} \tag{3.15}$$

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 $X_1, X_2, \ldots, X_v$ or $\pi_1 f_1, \pi_2 f_2, \pi_v f_v$,

Table 3.2: Confusion Matrix for the Classification Process

		Predicted	
		Yes	No
Actual	Yes	TP	FP
	No	FP	TN

TP is True positive: Observation is predicted positive and is actually positive.
FP is False positive: Observation is predicted positive and is actually negative,
TN is True Negative : Observation is predicted negative and is actually negative.
FN is False Negative: Observation is predicted negative and is actually positive.

(Hand, 1982), 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 optimization 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 the developed Discriminant Rule and 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 are 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 (Silverman, 2018) and (Hand, 1982). 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.

3.5 Comparisons Between the Developed Nonparametric Discriminant Function with LDA and QDA

In this section a description of methodology used to compare the developed Nonparametric Kernel Discriminant Function is given;

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

with Linear Discriminant Function;

$$x \ is \ allocated \ to \ group \ j_0 \ if j_0 = arg \max_{\{j \in 1,\ldots,v\}} \log(\pi_j) - \frac{1}{2}(x - \mu_j)^T \Sigma^{-1}(x - \mu_j)$$

Details of these functions are in sections 3.2 and 3.2.3. This comparison is through a simulation study in which some density functions were first considered, from which data was generated using Monte-Carlo Methods. The densities used in this thesis for simulation are adopted from the work done by (Duong, 2004). The generated dataset was divided into two mutually disjoint but exhaustive sets; a training and a testing set. The Training dataset was used to calibrate various Discriminant Functions. The Test dataset, was then subjected to the Calibrated Discriminant Functions to validate it.

Various replications were conducted, and for each iteration the Misclassification Rate was noted. This was done for various Discriminant Functions and the respective results finally averaged. The Discriminant Function with the smallest Misclassification rate was considered to be the best one. In the subsection 3.5.1, the Algorithm that was used to implement the simulation is given.

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.

Therefore, 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 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.

3.5.1 Algorithm for Proposed Kernel Discriminant Analysis

The algorithm for the proposed kernel discriminant analysis is given in this section. 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 $X_j = \left\{X_{j1}, X_{j2}, \ldots, X_{jn_j}\right\},\ j = 1, 2, \ldots, v$, compute a kernel density estimate

$$\hat{f}\left(x; H_j\right) = n_j^{-1} \sum_{i=1}^{n_j} K_{H_j}\left(x - X_{ji}\right) \tag{3.16}$$

   We can use any sensible bandwidth selector H_j

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

3. a. Allocate test data points $Y_1, Y_2, \ldots, Y_m$ according to KDR/Equation (3.4) or

   b. Allocate all points x from the sample space according to KDR/Equation (3.4).

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

$$\hat{MR} = 1 - m^{-1} \sum_{k=1}^{v} 1 \left\{ Y_K \ is \ classified \ correctly \ using \ KDR. \right\} \quad (3.17)$$

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

$$M\hat{R}_{CV} = 1 - n^{-1} \sum_{j=1}^{v} \sum_{i=1}^{n_j} 1 \left\{ X_{ji} \ is \ classified \ correctly \ using \ KDR_{ji}. \right\} \quad (3.18)$$

where 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

$$\hat{f}_{j,-i}(x; H_j) = (n_j - 1)^{-1} \sum_{i'=1, i' \neq 1}^{n_j} K_{H_{j,-i}}\left(x - X_{ji'}\right) \quad (3.19)$$

That is, we repeat step 3 to classify all X_{ji} using KDR_{ji}.

3.6 Classifying Stateless Communities

To use the Nonparametric Kernel Discriminant Function in a practical situation, data from the Kenya National Bureau of Statistics (KNBS) obtained from the 2009 Kenya Population and Housing Census and a report of a Survey on Pemba Stateless Community that was conducted in 2015 were used. Data from the said Census consists of tribes living in the Coastal Region of Kenya especially Kilifi County where majority of the Pemba Community lives. Various characteristics associated with these tribes such as Education level, Religion, Housing building materials (Housing Materials for the Floors, Walls and Roof), Waste Disposal, Source of Water and Employment Status, were considered.

The study aimed to identify one of these communities that have similar characteristics to those of Pemba. This information was used to classify the Pemba community which has been stateless for a long time. Due to the challenges of insufficient data in the database regarding Pemba Community, the only information available for use was based on the characteristics such as level of education and employment. The developed Nonparametric Kernel Discriminant Function, Linear Discriminant Function and the Quadratic Discriminant Function were applied to the training data so as to calibrate the said three functions.

Thereafter, the calibrated Functions were used to classify the test data from which the Pemba Community was fitted into particular local Neighboring one. The Misclassification rates were then obtained for each of the three functions in order to measure how accurate the classification function that classified the Pemba Community.

CHAPTER FOUR
RESULTS AND DISCUSSIONS

4.1 Introduction

The problem being addressed in this thesis falls into the area of survey sampling as follows; there exists a parent population with heterogeneous characteristics from which snowballing or related Discrimination sampling methods ought to be used to select members with a particular required characteristics thereby forming a sample. In particular, the study looks at already existing communities whose characteristics are known, but within which there are people who have been assimilated into the said communities yet should be in a stand alone formation.

4.2 Simulation Study

Sometimes in survey sampling, we do not usually observe all the survey information. That is, the survey variable is not observable for all the population units. Auxiliary variable X is often used to estimate the unobserved survey variables. One way of overcoming the above problem is the super population approach in which the working model relating the auxiliary variables to the response variable is assumed.

The following discriminant analysers were considered in this study.

i) Linear Discriminant (LD)

ii) Quadratic Discriminant (QD)

iii) Kernel Discriminant with 2-stage AMSE diagonal bandwidth matrices (KDD2)

iv) Kernel Discriminant with 2-stage SAMSE full bandwidth matrices (KDS2)

v) Kernel Discriminant with 1-stage SCV full bandwidth matrices (KDSC)

The bandwidths were generated from the data using various methods given as 2- stage AMSE, 2-stage SAMSE and 1-stage SCV, among others as described above. The R code for Kernel Discriminant Functions 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.

For a reasonable empirical study, the Normal Mixture Densities in (Duong, 2004) were used for the simulation study. These are the same densities that were earlier used in the empirical study by (Baudat and Anouar, 2000) whose study is comparable to this current one. More to this the density D contains fairly distinct components discriminant analyser and is expected to perform well here.

Density E has three components of various shapes and sizes, therefore, is more challenging case than density D. Density K is a pair of bimodal normal mixtures, with alternating modes. Density L is a large mode separating a bimodal density with narrower modes. For these two latter densities it is expected that, the linear and quadratic discriminant analysers to perform poorly since it is difficult to distinguish the different components using only linear or quadratic cuts.

Alternatively, densities K and L are highly Non-normal so the assumptions of Normality for the Parametric methods are invalid. Thus it is expected that the Kernel methods will demonstrate their efficiency here. The formulas for these target densities are provided in Table 4.1. Data was simulated from these densities for 500 trials using training Sample sizes $n = 100$ and $n = 1000$ and test data Sample size $k = 1000$. Different sample sizes were taken to check if they have any effect on the performance of the estimator. K is the number of iterations made to allow computing the average of the estimator with the aim of removing any possible biases.

Table 4.1: Formulas for Target Densities D, E, K & L

Target Density	Formular
D	$\pi_1 = \frac{1}{2}\ f_1 \sim N\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} \frac{4}{9} & \frac{14}{45} \\ \frac{14}{45} & \frac{4}{9} \end{bmatrix}\right);\ \pi_2 = \frac{1}{2},\ f_2 = N\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{4}{9} & 0 \\ 0 & \frac{4}{9} \end{bmatrix}\right)$
E	$\pi_1 = \frac{3}{7}\ f_1 \sim N\left(\begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{9}{25} & \frac{63}{250} \\ \frac{63}{250} & \frac{49}{100} \end{bmatrix}\right);\ \pi_2 = \frac{3}{7},\ f_2 = N\left(\begin{bmatrix} 1 \\ \frac{2}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{9}{25} & 0 \\ 0 & \frac{49}{100} \end{bmatrix}\right)$ $\pi_3 = \frac{1}{7}\ f_3 \sim N\left(\begin{bmatrix} 1 \\ -\frac{2}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{9}{25} & 0 \\ 0 & \frac{49}{100} \end{bmatrix}\right)$
K	$\pi_1 = \frac{1}{2}\ f_1 \sim \frac{1}{2}N\left(\begin{bmatrix} -\frac{3}{2} \\ -\frac{3}{2} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{4}{5} \end{bmatrix}\right) + \frac{1}{2}N\left(\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{4}{5} \end{bmatrix}\right);$ $\pi_2 = \frac{1}{2}\ f_2 \sim \frac{1}{2}N\left(\begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{4}{5} \end{bmatrix}\right) + \frac{1}{2}N\left(\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{4}{5} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{4}{5} \end{bmatrix}\right);$
L	$\pi_1 = \frac{1}{3}\ f_1 \sim \frac{1}{2}N\left(\begin{bmatrix} -\frac{3}{2} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{3}{10} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{10} \end{bmatrix}\right) + \frac{1}{2}N\left(\begin{bmatrix} \frac{3}{2} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{3}{10} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{10} \end{bmatrix}\right);$ $\pi_2 = \frac{2}{3}\ f_2 \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & 1 \end{bmatrix}\right)$

4.3 Misclassification Rates Using Simulated Data

The average and standard deviation of misclassification rates are in Table 4.2. From this table, for density D and E, the LD performed poorly compared to QD and the kernel discriminant analysers. For density K, our expectations are confirmed: KDD2, KDS2, KDSC all outperform the linear and quadratic counterparts. For density L, the advantage of the kernel methods over the linear method is maintained while it is reduced compared to the quadratic method.

The improved performance of the Kernel Discriminant Functions for the latter two densities is apparent for both sample sizes. Moreover, even with the increased burden of selecting an increased number of bandwidths which comprise the bandwidth matrix, the full matrix selectors overall produce smaller standard deviations. The differences between the diagonal matrix KDD2 and the full matrix KDSC and KDS2 are more subtle than the differences between the kernel methods and the parametric methods. We can see that both full bandwidth matrix methods KDS2 and KDSC in the majority of cases considered here have lower mean misclassification rates than KDD2.

Table 4.2: Misclassification Rates for Discriminant Functions

Target Density		Misclasifcation Rate				
		KDD2	KDS2	KDSC	LD	QD
n=100, k=1000						
D	mean	0.0793	0.00600	0.0050	0.0101	0.0050
	SD	0.0013	0.0013	0.0013	0.0042	0.0014
E	mean	0.00798	0.0719	0.0798	0.0703	0.0701
	SD	0.0120	0.0110	0.0071	0.0082	0.0081
K	mean	0.3008	0.2810	0.17989	0.6125	0.5998
	SD	0.0152	0.01320	0.0128	0.0396	0.0324
L	mean	0.1615	0.1517	0.1607	0.4010	0.2115
	SD	0.0156	0.0124	0.0146	0.0167	0.0189
n=1000, k=1000						
D	mean	0.0048	0.0050	0.0044	0.0059	0.039
	SD	0.0015	0.0015	0.0013	0.0014	0.0016
E	mean	0.0499	0.0499	0.0489	0.0509	0.0498
	SD	0.00146	0.0014	0.0014	0.0016	0.0013
K	mean	0.0498	0.0501	0.0498	0.0498	0.0497
	SD	0.0015	0.0016	0.0015	0.0014	0.0013
L	mean	0.4998	0.4996	0.4500	0.5189	0.4984
	SD	0.0154	0.0155	0.0154	0.0121	0.0164

4.4 Data Obtained From the 2009 Census

For a real application, we are using data from obtained from the 2009 Census. The data consist of tribes living in the Coastal Region of Kenya especially Kilifi County. The study considered using various characteristics associated to this community such as Education Level, Religion, Building Material, Waste Disposal, Source of Water and Employment Status.

The study aimed to classify these communities using the characteristics observed amongst them and also obtain the misclassification error which is the error that the community is classified in the wrong group. In addition, the study aimed at using this characteristics to classify the Pemba community. The results obtained will be useful to policy makers to consider integrating the Pemba people into the identified local community(ies). The results will further help to inform on the classification decision on any emerging tribe in the coastal region whose citizenship is not known but possess similar characteristics.

Whereas the Census Data had all the desired variables, the Pemba Community survey had a limited number. The only data available for use is based on the characteristics such as level of education and employment. We apply non parametric Discriminant Functions and compare their performance with the Parametric methods.

There are about 10 communities in Kilifi County with a population of about 1.02 million people which are neighboring the Pemba community with an estimated population of over $2,000$ people. Although some Pemba were issued with IDs in Kenya, most of them 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 people were asked to leave the country but they opted to 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 cannot even enjoy the available banking services.

To analyse this data and perform a classification, a sample of $3,000$ observations was taken using stratified simple random sampling where each the ten tribes was treated as a stratum. The proportional allocation technique was used to obtain a sample from each tribe to ensure equal representation in the study.

The sample data was then divided in two parts of 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.

4.5 Misclassification Rates for Stateless Communities in Kenya

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

Table 4.3: Misclassification Rates for Various Discriminant Functions using the 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 4.3 presents the Misclassification rates for various Discriminant classifiers. From these findings, it was observed that the Kernel Discriminant Classifiers KDDS2, KDS2 and KDSC had lower rates of Misclassification of 8.13%, 7.5%, and 8.75% respectively when used to classify each community into the correct community compared to their Parametric counterparts, Linear and Quadratic Discriminant Function with Misclassification Rates of 36.25% and 15.65% respectively. Thus, the Kernel classifier with appropriate choice of the bandwidth is recommended and considered to be more efficient for classifying purposes compared to the Parametric Functions.

Table 4.4: Misclassification Rates for each group for Various Discriminant Functions using the Training Dataset as a Test Data

Ethnic Group	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 4.4 presents the results for the misclassification rates within the groups/ communities. From the results, it was observed that, even within the groups, the Nonparametric Kernel discriminant classifiers exhibit lower Misclassification rates compared to their Parametric counterparts hence making them more efficient and reliable when classifying stateless communities. In the second analysis, the training data was used to train the model and an independent data used as a test data to see how the model performs.

Table 4.5: Misclassification Rates for Various Discriminant Functions using an Independent Test Dataset

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

From the results in Table 4.5, 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.

Table **??** that follows, gives these performance measures for each community based on different classifiers in determining how correctly that community was classified. It can be observed that the Kernel discriminant classifiers outperforms parametric classifiers when the appropriate bandwidth matrix is chosen as they show high values of precision, sensitivity, specificity and F1 across the Tribes/ethnic Groups.

Table 4.6: Classification Performance of Four Classification Models based on the Stateless Communities' Dataset

		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

4.6 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 Kenyan and be issued with the National Identification Number to enable them access Government Services without discrimination. Such activities include access to some basic rights and services like acquisition of birth certificates, education, formal employment, financial services. 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 Tables 4.7, 4.9 , 4.11 and 4.13 present the Confusion Matrix for the classification of the communities using the Kernel Discriminant Function KDD2, KDSC and the Quadratic and Linear discriminant Functions respectively. From these results, 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.

From the results, it can be seen that 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 the strongest link or to Rabai community.

Table 4.7: The Confusion Matrix of the Communities in Kilifi County classified based on KDD2 classifier

		Predicted											
		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
Observed	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 4.7 present the Confusion Matrix for the classification of communities using the Kernel Discriminant Function KDD2. From these results, the KDD2, 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 as Rabai. From this, it can be observed that, based on certain similarities that exist in these communities, the Pemba can be assimilated into Giriama because they seem to have the strongest links or to Rabai.

Table 4.8: The Confusion Matrix of the Proportion of the Communities being classified correctly into a particular community based on KDD2 classifier

		Predicted										
		Bajuni	Boni	Digo	Duruma	Giriama	Jibana	Kambe	Pemba	Rabai	Ribe	Waata
Observed	Bajuni	0.00000	0.00000	0.00000	0.00000	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	0.00000	0.00000	0.00000
	Digo	0.00000	0.00000	0.00000	0.00000	1.0000	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	0.98402	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	0.00000	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	0.0000	0.63504	0.15328	0.00000	0.00000
	Rabai	0.00000	0.00000	0.00000	0.0000	0.35000	0.00000	0.0000	0.28750	0.36250	0.00000	0.00000
	Ribe	0.00000	0.00000	0.00000	0.0000	0.97059	0.00000	0.0000	0.00000	0.00000	0.00000	0.02941
	Waata	0.00000	0.00000	0.00000	0.00000	0.97297	0.00000	0.0000	0.02703	0.00000	0.00000	0.00000

Table 4.8 has the Proportion of communities being classified correctly into a particular neigbouring community using the Kernel Discriminant Classification Function KDD2. From these results, it can be observed that, there is 21.2% chance of correctly classifying Pemba as Giriama and 15.2% chance of classifying them correctly as Rabai. The chance of classifying them into other communities is zero since they do not share similar characteristics.

Table 4.9: The Confusion Matrix of the Communities in Kilifi County classified based on KDSC classifier

		Predicted											
		Bajuni	Boni	Digo	Duruma	Giriama	Jibana	Kambe	Pemba	Rabai	Ribe	Waata	Total
Observed	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

The results in Table 4.9, present the Confusion Matrix for the classification of the communities using the Kernel Discriminant Function KDSC. From these results, majority The KDSC function classified 29 people as Giriama, 88 as Pemba and 20 as Rabai. From this finding it can be observed that, based on certain similarities that exist in this communities, the Pemba Community can be classified as Giriama and also Rabai Community.

Table 4.10: The Confusion Matrix of the Proportion of the Communities being classified Correctly into a Particular Community based on KDSC Classifier

		Predicted										
		Bajuni	Boni	Digo	Duruma	Giriama	Jibana	Kambe	Pemba	Rabai	Ribe	Waata
Observed	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

The results in Table 4.10 present the Proportion of the communities being classified correctly into a particular community using the Kernel Discriminant Classification Function KDSC. From these results, it can be observed that, there is 21.2% chance of correctly classifying Pemba as Giriama and 14.6% chance of classifying them correctly as Rabai. From these results, it can also be seen that, the chance of Giriama being classified as Pemba Community was very small with probability of 2.28%. The chance of classifying them into other communities is zero since they do not share similar characteristics.

Table 4.11: The Confusion Matrix of the Communities in Kilifi County classified based on QDA Classifier

		Predicted											
		Bajuni	Boni	Digo	Duruma	Giriama	Jibana	Kambe	Pemba	Rabai	Ribe	Waata	Total
Observed	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
	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

The results in Table 4.11 present the confusion matrix for the classification of the communities using Quadratic Function. From these results, QDA classification function classified majority 20 as Giriama, 79 as Pemba and 20 as Rabai. From this finding it can be observed that, based on certain similarities that exist in this communities, the Pemba Community can be classified as Giriama because they seem to have the strongest link or to Rabai community.

Table 4.12: The Confusion Matrix of the Proportion of the Communities being classified correctly into a particular community based on QDA Classifier

		Predicted										
		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	0.02940	0.00000
	Digo	0.02780	0.00000	0.00000	0.02780	0.86110	0.02780	0.00000	0.00000	0.02780	0.00000	0.02780
	Duruma	0.00000	0.00000	0.01450	0.02900	0.84060	0.01450	0.00000	0.02900	0.02900	0.00000	0.04350
Observed	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	0.27270	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	0.28750	0.00000	0.01250
	Ribe	0.02940	0.0294	0.00000	0.02940	0.76470	0.05880	0.00000	0.02940	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

The results in Table 4.12 present the Proportion of the communities being classified correctly into a particular community using the QDA classification function. From these results, it can be observed that, there is 21.2% chance of correctly classifying Pemba as Giriama and Rabai was 27.1%. The chance of classifying them into other communities is there but very small making the Giriama and Rabai the dominant community where can be integrated.

Table 4.13: The Confusion Matrix of the communities in Kilifi County classified based on LDA classifier

		Predicted											
		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
Observed	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

The results in Table 4.13 present the confusion matrix for the classification of the communities using the Linear Discriminant Classification Function. From these results, the LDA Classification Function 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 the strongest link or to Rabai Community.

Table 4.14: The Confusion Matrix of the Proportion of the communities being classified correctly into a particular community based on LDA classifier

		Predicted										
		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
Observed	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

The results in Table 4.14 present the Proportion of the communities being classified correctly into a particular community using the LDA classification function. From these results, it can be observed that, chance of correctly classifying Pemba as Giriama and Rabai is 16.1% and 13.9% respectively. The chance of classifying them into other communities is there but very small making the Giriama and Rabai the dominant community where can be integrated.

CHAPTER FIVE
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

In this final Chapter, a brief summary of key results from this thesis are presented. In doing so, a summary that expounds on what the thesis alluded to at the abstract stage has been provided followed by conclusions on each specific objective and ensuing recommendations guided by the theoretical and empirical analysis in this thesis. From all these, it is finally concluded that stateless persons can be identified and correctly assimilated into neighboring communities in Kenya.

5.2 Summary

In achieving the set out objectives, this thesis has demonstrated that it is possible to replace the parametric function within the general Bayes Discriminant Rule with an optimally chosen Kernel function thereby yielding a robust Nonparametric Discriminant Function. The performance of this Nonparametric Discriminant Function was found to be better when compared with Linear and Quadratic Discriminant Funtions.

This is therefore a milestone when used in context of Survey sampling, treating the existing communities as a parent population with heterogeneous characteristics from which a homogeneous sample is identified and correctly classified.

In the field of classification, especially the use of Nonparametric Discriminant Function had not been explored much in literature especially in the application when classifying the disadvantaged group like the stateless people. On the other hand, much has been discussed and application made in the Parametric Discriminating Functions. The challenge with these Functions is that they fail to capture the nonlinearity that occurs mainly in the real life datasets.

Further, most of the real life datasets do not assume the normality assumptions that are imposed by the parametric discriminating functions such as Linear and Quadratic functions. Therefore, this study has improved on existing literature and for application purposes, another alternative discriminant classification method that can be used where parametric functions fail has been brought forth.

During empirical analyses, some important variables were added into the 2019 Kenya Population and Housing Census data. This was a lesson from the previous census and so a way of capturing statelessness in Kenya was explored a decision was made to include statelessness as an option under ethnicity/nationality. However, it still lumped the stateless groups without specifically breaking down each of them.

Furthermore, the Census results still showed low numbers than expected because some stateless persons opted not to declare their stateless situation possibly for fear of victimization by the authorities. In this Thesis, it has been possible to tackle the issue of statelessness and it is hoped that going forward, Censuses will capture information that may be used to classify any persons who during the Census night may be recorded as stateless.

The Global Sustainable Development Goals (SDGs) or the Agenda 2030 as are known, are founded on the principle of "leaving no one behind (LNOB)". It represents the unequivocal commitment of all United Nation member states for example, to end discrimination and exclusion, and reduce the inequalities and vulnerabilities that leave some people behind. This Thesis has attempted to create an avenue for achieving this goal so as not to undermine the potential of individuals and of humanity as a whole. One of the causes of people being left behind has been persistent forms of discrimination, including failure to recognize such vulnerable people, which leaves individuals, families and whole communities marginalized and excluded. The results of this Thesis may be used as a means of addressing such challenges each time they occur anywhere around the world.

The work whose results are herein summarized has showed that in as much as

SDGs stress on leaving no one behind, these groups of people may be left further behind if they are neither given citizenship nor fully assimilated into the surrounding communities. The thesis has presented a method that may be employed to Robustly assimilate the stateless people into surrounding communities.

With majority of stateless communities in Kenya and around the world remaining undisclosed, the findings of this study may be used by the Government of Kenya as a guide on which local community could be used to integrate stateless communities who have lived in Kenya for long. This will help in making service delivery to such stateless communities and go a long way in achieving the proposals by the UNHCR in finding a solution on how to deal with stateless people residing in each particular country.

5.3 Conclusions

The main objective of this study was to classify the stateless communities using a Robust Nonparametric Kernel Discriminant Function. In this thesis, the Robust Nonparametric Kernel Discriminant Function was then used to find which neighboring local community in Kilifi County that the Pemba community can be integrated into to pave way for their recognition as Kenyans. This will eventually enable them acquire national identification cards and hence gain access to services from the Government without discrimination, just like any other Kenyans.

In particular, the first objective of this thesis was to develop a Nonparametric Discriminant Function by replacing the Parametric Discriminant Function in the Bayes Discriminant Rule with a Nonparametric Kernel Discriminant Function. This study developed Nonparametric Discriminant Function as given in equation 3.4 in section 3.2.3.

The second objective of this thesis was to estimate the Classification Rates of the developed Nonparametric Kernel Discriminant Function as a measure of its Robustness. The Classification Rates derived from Misclassification rates of the developed Nonparametric Kernel Discriminant Function were estimated as provided in equation 3.15 in section 3.4. The classification rates follows from the traditional Bayes Classification rates.

The third objective of this thesis was to compare the developed Nonparametric Kernel Discriminant Function with the Linear Discriminant Function and the Quadratic Discriminant Function through a simulation study. When applied to simulated data the findings indicate that the developed Nonparametric Kernel Discriminant Function has the lowest values of Misclassification rates compared to other existing Discriminant Functions such Linear Discriminant Function and the Quadratic Discriminant Function which are frequently used in real life.

As evidenced from the analysis of Misclassification rates presented in Table 4.2, it was possible to significantly reduce the misclassification rate thereby increasing the classification precision which depicts the developed discriminant function as being more Robust. The Misclassification rates indicate that the developed Nonparametric Kernel Discriminant Function is superior to the other Discriminant Function in all the densities used and various bandwidths.

Finally, the Fourth objective of this thesis was to use the developed Nonparametric Kernel Discriminant Function in classifying stateless Communities in Kenya. When applied to data obtained from the 2009 Kenya Population and Housing Census and a report on Pemba stateless Community conducted in 2015, the findings indicate that the developed Kernel discrimination can be relied upon in the classification of stateless communities in Kenya to find which communities they can be integrated into among the local communities.

As evidenced from the analysis of the Misclassification rate when classifying the Pemba Community presented in Tables 4.3, 4.4 and 4.5, the Nonparametric Kernel Discriminant Function exhibited lower misclassification rates hence making it more efficient and reliable when classifying stateless communities. Additionally, the Confusion Matrices for the classification of communities using the Kernel Discriminant Function KDD2, KDSC and the Quadratic and Linear Discriminant Functions presented in Tables 4.7, 4.9, 4.11 and 4.13.

The developed Kernel Discriminant Function classifies majority 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 Linear Discriminant Functions classified the Pemba people with 22 being classified as Giriama, 76 as true Pemba and 19 as the Rabai. Therefore, based on certain similarities that exists in this communities, it can be concluded that the Pemba community can be classified as Giriama or Rabai community because they seem to have the strongest link.

Additionally from the results, the following observations and conclusions have been made;

1. 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.

2. Both full bandwidth matrix methods KDS2 and KDSC in the majority of cases considered here have lower mean misclassification rates than KDD2.

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

4. The Nonparametric Kernel Discriminant Classifiers; 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 as Rabai. The KDSC classifier classified 29 people as Giriama, 88 as Pemba and 20 as Rabai.

5. 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.

6. Based on certain similarities in characteristics that exist in these communities that surround the Pemba, they can be classified as Giriama in which they seem to have the strongest link. The next strong link is with the Rabai community.

5.4 Recommendations

The study recommends use of a Kernel Discriminant Function to classify the stateless communities in Kenya and in particular, from our empirical analysis based on data from Kenya National Bureau of Statistics (KNBS) obtained from the 2009 Kenya Population and Housing Census and a report on Pemba stateless Community conducted in 2015, the study recommends that Pemba community be integrated into Giriama.

The approach that has been employed in this thesis can be extended to all the stateless communities in Kenya. Statelessness being a worldwide problem, the methodology used in this study can also be extended to all the cases of statelessness around the world.

In doing what the thesis has demonstrated for the Pemba Community across our world, it will go a long way in achieving the 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 through Surveys so as to allow more analyses and improve the efficiency of the results obtained.

Finally, the study recommends other classifications techniques which can handle the high dimensional spaces such as Neural Networks to be considered in the future studies so as to see if efficiency of classification can be improved.

REFERENCES

Abuya (2010). Out of the shadows towards ensuring the rights of stateless persons and persons and risk of statelessness. *Oxford Brookes University, 40.*

Adams, N. M. and Hand, D. J. (1999). Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognition*, 32(7):1139–1147.

Aizerman, M., Braverman, E., and Rozonoer, L. (1964). Probability problem of pattern recognition learning and potential functions method. *Avtomat. i Telemekh*, 25(9):1307–1323.

Akaho, S. (2006). A kernel method for canonical correlation analysis. *arXiv preprint cs/0609071.*

Albarazi, Z. (2016). Syrian refugee or stateless refugee: The challenges of statelessness in exile. *Middle East Centre Blog.*

Alizadeh, M. and Ebadzadeh, M. M. (2011). Kernel evolution for support vector classification. In *2011 IEEE workshop on evolving and adaptive intelligent systems (EAIS)*, pages 93–99. IEEE.

Almuallim, H. and Dietterich, T. G. (1994). Learning boolean concepts in the presence of many irrelevant features. *Artificial intelligence*, 69(1-2):279–305.

AmnestyInternational (2015). Dominican republic: 'without paper, i am no one': Stateless people in the dominican republic. *Amnesty International.*

AmnestyInternational (2017). The world's stateless children. *Institute on Statelessness and Inclusion.*

Assembly, U. G. (2014). Universal declaration of human rights, 1948, article 15. *See www. ohchr. org/EN/UDHR/Documents/UDHR_ Translations/eng. pdf, last accessed on*, 23.

AU, A. U. (2018). General comment on article 6 of the african charter on the rights and welfare of the child.

Aye, P. (2021). Predicting students performance in the university system using discriminant analysis method. *Multidisciplinary International Journal of Research and Development*, 1(02).

Balamurali, S. and Kalyanasundaram, M. (2011). An investigation of the effects of misclassification errors on the analysis of means. *Tamsui Oxford Journal of Information and Mathematical Sciences*, 27(2):117–136.

Baudat, G. and Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural computation*, 12(10):2385–2404.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83(3):377.

Bloch, A. and Donà, G. (2019). *Forced Migration*. Routledge.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

Bosire, C. (2017). Statelessness in a state? *Available at SSRN 3079430*.

Brause, R. W., Langsdorf, T. S., and Hepp, H.-M. (1999). *Credit card fraud detection by adaptive neural data mining*. Univ.-Bibliothek Frankfurt am Main.

Breckenridge, J. N. (2000). Validating cluster analysis: Consistent replication and symmetry. *Multivariate Behavioral Research*, 35(2):261–285.

Chhikara, R. S. and McKeon, J. (1984). Linear discriminant analysis with misallocation in training samples. *Journal of the American Statistical Association*, 79(388):899–906.

Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.

Cole, P. (2019). Taking statelessness seriously. *Statelessness & Citizenship Rev.*, 1:161.

Coomans, F. (2005). Reviewing implementation of social and economic rights: an assessment of the 'reasonableness' test as developed by the south african constitutional court. *Heidelberg Journal of International Law*, 65:167–96.

Diaf, A., Boufama, B., and Benlamri, R. (2012). A compound eigenspace for recognizing directed human activities. In *International Conference Image Analysis and Recognition*, pages 122–129. Springer.

Diaf, A., Boufama, B., and Benlamri, R. (2013). Non-parametric fisher's discriminant analysis with kernels for data classification. *Pattern recognition letters*, 34(5):552–558.

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, 27(3):326–327.

Dietterich, T. G., Hild, H., and Bakiri, G. (1995). A comparison of id3 and backpropagation for english text-to-speech mapping. *Machine Learning*, 18(1):51–80.

DiStefano, C. and Morgan, G. (2011). Examining classification criteria: A comparison of three cut score methods. *Psychological Assessment*, 23(2):354.

Du, C., Zhou, S.-l., Sun, J.-x., Sun, H., and Wang, L.-l. (2013). Discriminant embedding by sparse representation and nonparametric discriminant analysis for face recognition. *Journal of Central South University*, 20(12):3564–3572.

Duong, T. (2004). *Bandwidth selectors for multivariate kernel density estimation.* University of Western Australia Perth.

Erimafa, J., Iduseri, A., and Edokpa, I. (2009). Application of discriminant analysis to predict the class of degree for graduating students in a university system. *International Journal of Physical Sciences*, 4(1):16–21.

Evas, T. and Väljataga, A. (2016). Mapping statelessness in estonia. *UN High Commissioner for Refugees (UNHCR).*

Fan, Z., Xu, Y., and Zhang, D. (2011). Local linear discriminant analysis framework using sample neighbors. *IEEE Transactions on Neural Networks*, 22(7):1119–1132.

Fernandez, G. C. (2002). Discriminant analysis, a powerful classification technique in data mining. In *Proceedings of the SAS users international conference*, pages 247–256. Citeseer.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188.

Flury, B. (2013). *A first course in multivariate statistics.* Springer Science & Business Media.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American statistical association*, 84(405):165–175.

Fukunaga, K. and Mantock, J. (1983). Nonparametric discriminant analysis. *IEEE Transactions on pattern analysis and machine intelligence*, (6):671–678.

Garden, A. (1981). Classification: methods for the exploratory analysis of multivariate data.

Garreton, R. (1997). Report on the situation of human rights in zaire, prepared bythe special rapporteur, mr. roberto garreton, in accordance with commission resolution 1996/77. *Refugee Survey Quarterly*, 16(1):42–95.

Ghassabeh, Y. A., Rudzicz, F., and Moghaddam, H. A. (2015). Fast incremental lda feature extraction. *Pattern Recognition*, 48(6):1999–2012.

Gu, X., Liu, C., Wang, S., and Zhao, C. (2015). Feature extraction using adaptive slow feature discriminant analysis. *Neurocomputing*, 154:139–148.

Hall, P. and Wand, M. P. (1988). On nonparametric discrimination using density differences. *Biometrika*, 75(3):541–547.

Hand, D. J. (1982). Kernel discriminant analysis. *JOHN WILEY & SONS, INC., ONE WILEY DR., SOMERSET, N. J. 08873, 1982, 264*.

Höfler, M. (2005). The effect of misclassification on the estimation of association: a review. *International journal of methods in psychiatric research*, 14(2):92–101.

Holden, J. E. and Kelley, K. (2010). The effects of initially misclassified data on the effectiveness of discriminant function analysis and finite mixture modeling. *Educational and psychological measurement*, 70(1):36–55.

Howley, T. and Madden, M. G. (2005). The genetic kernel support vector machine: Description and evaluation. *Artificial intelligence review*, 24(3-4):379–395.

Information, C., Policy Unit, A., and Directorate, A. P. (2022). Uk home office immigration and nationality directorate country report - turkey. *United Kingdom: Home Office.*

Jann, B. (2007). Univariate kernel density estimation. *Boston College Department of Economics, Statistical Software Component*, (S456410).

Jarchi, D. and Boostani, R. (2006). A new weighted lda method in comparison to some versions of lda. *Proceedings of Word Academy of Science, Engineering and Technology*, 12(18):233–238.

Johnson, R. A., Wichern, D. W., et al. (2014). *Applied multivariate statistical analysis*, volume 6. Pearson London, UK:.

KNBS (2009). Kenya population and housing census.

KNBS (2021). Economic survey. *Kenya National Bureau of Statistics*.

Ksantini, R. and Boufama, B. (2012). Combining partially global and local characteristics for improved classification. *International Journal of Machine Learning and Cybernetics*, 3(2):119–131.

Ksantini, R., Boufama, B., and Ahmad, I. S. (2011). A new ksvm+ kfd model for improved classification and face recognition. *Journal of Multimedia*, 6(1).

Lachenbruch, P. A. (1966). Discriminant analysis when the initial samples are misclassified. *Technometrics*, 8(4):657–662.

Lachenbruch, P. A. (1974). Discriminant analysis when the initial samples are misclassified ii: Non-random misclassification models. *Technometrics*, 16(3):419–424.

LawyersforHumanRights (2014). *Promoting Citizenship and Preventing Statelessness in South Africa: A Practitioner's Guide*. Pretoria University Law Press (PULP).

Lei, P.-W. and Koehly, L. M. (2003). Linear discriminant analysis versus logistic regression: A comparison of classification errors in the two-group case. *The Journal of Experimental Education*, 72(1):25–49.

Li, J.-B., Sun, W.-H., Wang, Y.-H., and Tang, L.-L. (2013). 3d model classification based on nonparametric discriminant analysis with kernels. *Neural Computing and Applications*, 22(3):771–781.

Li, Q., Wei, F., and Zhou, S. (2017). Local kernel nonparametric discriminant analysis for adaptive extraction of complex structures. *Open Physics*, 15(1):270–279.

Loog, M., Duin, R. P. W., and Haeb-Umbach, R. (2001). Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE transactions on pattern analysis and machine intelligence*, 23(7):762–766.

Lynch, M. and Ali, P. (2006). *Buried Alive: Stateless Kurds in Syria*. Universitäts-und Landesbibliothek Sachsen-Anhalt.

Manby, B. (2011). Statelessness in southern africa. In *Briefing paper for UNHCR Regional Conference on Statelessness in Southern Africa*.

Manby, B. (2018). Statelessness and citizenship in the east african community. *Available at SSRN 3783292*.

Mangasarian, O. L. (1965). Linear and nonlinear separation of patterns by linear programming. *Operations research*, 13(3):444–452.

Manly, M. (2012). Unhcr's mandate and activities to address statelessness in europe. *European Journal of Migration and Law*, 14(3):261–277.

Manly, M. and Van Waas, L. (2014). The state of statelessness research: A human rights imperative. *Tilburg Law Review*, 19(1-2):3–10.

McLachlan, G. (1972). Asymptotic results for discriminant analysis when the initial samples are misclassified. *Technometrics*, 14(2):415–422.

Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K.-R. (1999). Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pages 41–48. Ieee.

Milbrandt, J. (2011). Stateless. *Cardozo J. Int'l & Comp. L.*, 20:75.

Muimi, A. M. (2021). Generational statelessness and rights a case of children from the nubian community in kenya.

Onu, J. O. and John, O. O. (2016). An application of discriminant analysis to the classification of students on the basis of their academic performances. *Statistics*, 56:4.

Ozasa, K. (2008). The effect of misclassification on evaluating the effectiveness of influenza vaccines. *Vaccine*, 26(50):6462–6465.

Panigrahi, R. and Borah, S. (2018). Classification and analysis of facebook metrics dataset using supervised classifiers. *Social Network Analytics: Computational Research Methods and Techniques*, 1.

Persons, S. (2012). Guidelines to protect stateless persons from arbitrary detention: Introduction. *European Journal of Migration and Law*, 14:335–349.

Pflug, G. (1996). Kernel smoothing. monographs on statistics and applied probability-mp wand; mc jones. *Metrika*, 43:266–266.

Qiu, X. and Wu, L. (2006). Nearest neighbor discriminant analysis. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(08):1245–1259.

Roth, V. and Steinhage, V. (1999). Nonlinear discriminant analysis using kernel functions. In *NIPS*, volume 12, pages 568–574.

Schölkopf, B., Mika, S., Smola, A., Rätsch, G., and Müller, K.-R. (1998). Kernel pca pattern reconstruction via approximate pre-images. In *International Conference on Artificial Neural Networks*, pages 147–152. Springer.

Scott, D. W. (1991). Feasibility of multivariate density estimates. *Biometrika*, 78(1):197–205.

Scott, D. W. (1992). Multivariate density estimation: Theory, practice and visualisation. john willey and sons. *Inc., New York*.

Shawe-Taylor, J., Cristianini, N., et al. (2004). *Kernel methods for pattern analysis*. Cambridge university press.

Shi, Z. and Hu, J. (2012). Local linear discriminant analysis with composite kernel for face recognition. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–5. IEEE.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Simonoff, J. S. (1996). Further applications of smoothing. In *Smoothing Methods in Statistics*, pages 252–274. Springer.

Sireci, S. G., Robin, F., and Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education*, 12(3):301–325.

Southwick, K. and Lynch, M. (2013). Nationality rights for all: A progress report and global survey on statelessness. refugees international. *Refugees International (March 2009).*

Sun, M. and Xiong, M. (2003). A mathematical programming approach for gene selection and tissue classification. *Bioinformatics*, 19(10):1243–1251.

Sutton, N. L. J. (2018). Statelessness and the rights of children in kenya and south africa: A human rights perspective.

Thomas, A. U. and Pascal, B. N. (2013). A comparison of theperformance of students in pre-degree and universitymatriculation examination classes in a university system: Adiscriminant analysis approach. *International Journal of Adv.in Research and Technology, vol. 2, Issue 9.*

Thomas, L., Crook, J., and Edelman, D. (2017). *Credit scoring and its applications*. SIAM.

Tucker, J. (2021). The statelessness of refugees. In *Statelessness, governance, and the problem of citizenship*, pages 61–73. Manchester University Press.

UNCR, U. H. C. (2020). The impact of covid-19 on stateless populations: Policy recommendations and good practices. *UN High Commissioner for Refugees UNHCR.*

UNHCR (2014a). Background note on gender equality, nationality laws and statelessness. *http://www.refworld.org/docid/532075964.html.*

UNHCR (2014b). Global action plan to end statelessness 2014–2024.

UNHCR (2017). Representing stateless persons before u.s. immigration authorities: A legal practice resource from the united nations high commissioner for refugees. *UN High Commissioner for Refugees (UNHCR).*

UNHCR (2022). Good practices paper - action 1: Resolving existing major situations of statelessness. *UN High Commissioner for Refugees (UNHCR).*

Usoro, A. (2006). Discriminant analysis and its application to the classification of students on the basis of their academic performances. *J. Res. Phy. Sci*, 2(3):53–55.

Van Waas-Hayward, L. (2008). Nationality matters: statelessness under international law.

Van Wass, L. (2010). The situation of stateless persons in the middle east and north africa. *Geneva: United Nations High Commissioner for Refugees (UNHCR)[online] Available at http://www. refworld. org/docid/4cea28072. html [Accessed 4 June 2018].[Google Scholar]*.

Weissbrodt, D. (2000). Prevention of discrimination-the rights of non-citizens. *Final Report of the Special Rapporteur, Mr. David Weissbrodt, submitted in accordance with Sub-Commission decision*, 103.

Yu, H. and Yang, J. (2001). A direct lda algorithm for high-dimensional data—with application to face recognition. *Pattern recognition*, 34(10):2067–2070.

Zeng, Q. (2014). Weighted marginal discriminant analysis. *Neural Computing and Applications*, 24(3):503–511.

Zigler, E. and Phillips, L. (1961). Social competence and outcome in psychiatric disorder. *The Journal of Abnormal and Social Psychology*, 63(2):264.