MAKERERE UNIVERSITY

COLLEGE OF COMPUTING AND INFORMATION SCIENCES

NAME: NKATA JOSHUA LUYOMBYA

REG. NO: 2025/HD05/26365U

STUDENT NO: 2500726365

COURSE: MASTER OF SCIENCE IN COMPUTER SCIENCE

COURSE UNIT: MACHINE LEARNING

COURSE CODE: MCS 7103

PROJECT REPORT: RESOURCE ALLOCATION PREDICTION IN UGANDA

Introduction

This project develops a machine learning system to predict healthcare resource allocation needs in Uganda using demographic health survey data. The system employs a two-stage approach: first, predicting population pressure scores using regression, then classifying districts into risk categories for targeted resource allocation.

1. Problem Statement and Objectives

Resource allocation in developing countries faces significant challenges due to:

- Limited resources and competing demands
- Population growth and demographic pressures
- Varying health indicators across regions
- Need for data-driven decision making

Objectives

- Primary: Develop predictive models to identify districts with high resource allocation needs
- Secondary: Create interpretable scoring systems for policy makers
- Tertiary: Establish a framework for ongoing monitoring and prediction

2. Data Description and Preprocessing

Dataset Overview

- Source: DHS Quickstats Subnational Uganda Dataset
- Initial Size: 1,828 records across 30 variables
- Coverage : Multiple districts and survey years
- Data Type: Health demographic indicators and survey responses

Key Variables

- Location: District identifiers
- Temporal: Survey years
- Health Indicators: Fertility rates, mortality rates, contraceptive use
- Risk Factors: Malnutrition, vaccination coverage, delivery locations

Data Preprocessing Steps

- Column Standardization: Cleaned names(lowercase, underscores)
- Data Type Conversion: Converted numeric strings to proper types
- Missing value Treatment: Forward/backward filling within district groups
- Data Filtering: Removed invalid entries (e.g, “data+year”)
- Feature Engineering: Created composite scores and categories

3. Methodology

Composite Score Development

Population Pressure Score

Formula used: 0.5 x Fertility + 0.3 x (1/Mortality) + 0.2 x (1- Contraceptive_Use)

- Rationale: Higher fertility and lower contraceptive use indicate resource pressure
- Weights: Based on demographic impact on resource needs

Nutrition Index

Components: Average of child stunting, wasting, and underweight rates

- Purpose: Measures malnutrition burden requiring intervention

Health Service Demand

Components: Healthcare facility delivery rates and vaccination coverage

- Purpose: Indicates current healthcare system utilization

Resource Risk Score

Formula used: 0.6 x Pressure + 0.25 x Nutrition + 0.15 x Health_Demand

- Integration: Combines multiple health dimensions
- Output: Single metric for resource allocation priority

4. Machine learning Methodology

Two-Stage Approach

Stage 1: Pressure Score Prediction(Regression)

- Algorithm: Random Forest Regressor(200 estimators)
- Input Features: All numeric health indicators
- Target: Population pressure score
- Purpose: Predict continuous pressure values for new data

Stage 2: Risk Category Classification

- Algorithm: Random Forest Classifier (200 estimators)
- Input Features: Predicted pressure score, nutrition index, and health service demand
- Target: Risk categories (High/Medium/Low)
- Purpose: Actionable classification for resource allocation

5. Key Findings/Results

Regression Model – Predicting Pressure Score

```
MSE: 0.548941764206482
R^2: 0.907242786234734

Sample Predicted Pressure Scores:
indicator  pressure_score  predicted_pressure_score
0               -3.194958                 -3.171698
2               -5.274783                 -5.252041
4               -5.794444                 -5.689130
6               -4.234872                 -4.343496
8               -2.876104                 -2.954124
10              -2.745683                 -2.747832
12              -6.464737                 -6.293266
14              -5.684783                 -5.755888
16              -3.223684                 -3.257065
18              -5.165489                 -5.281331
```

- Shows how well the model predicts population pressure
- Helps determine reliability of Stage 2

Classification Model – Predicting Pressure

```
Resource Risk Classification Report:
              precision    recall  f1-score   support

   High Risk       1.00      0.80      0.89         5
    Low Risk       1.00      1.00      1.00         1
 Medium Risk       0.80      1.00      0.89         4

    accuracy                           0.90        10
   macro avg       0.93      0.93      0.93        10
weighted avg       0.92      0.90      0.90        10


Sample Risk Category Predictions:
   Actual Risk Predicted Risk
0  Medium Risk    Medium Risk
1  Medium Risk    Medium Risk
2  Medium Risk    Medium Risk
3    High Risk      High Risk
4  Medium Risk    Medium Risk
5    High Risk    Medium Risk
6    High Risk      High Risk
7    High Risk      High Risk
8    High Risk      High Risk
9     Low Risk       Low Risk
```

Highest Risk Districts

```
Top 5 Highest Risk Districts (by avg pressure score):
  Karamoja: 1.201
  West Nile: 0.408
  Northern: -0.170
  Eastern: -0.519
  Western: -0.997
```

6. **Recommendations**

- Deploying Prediction System: Implement models for ongoing resource allocation decisions
- Focusing High-Risk Districts: Prioritizing identified high-risk areas for immediate intervention
- Policy Integration: Incorporate predictive scores into official resource allocation criteria

7. Conclusion

This project successfully demonstrates the application of machine learning techniques to resource allocation challenges in Uganda. The two-stage predictive system provides both continuous pressure scores and discrete risk categories, enabling flexible decision-making for policy makers.