**Appendix: Project Report on World Statistics Dataset Analysis**

**Software outputs organized by techniques used**

1. **Classification tree for prediction of Development state of country**

Software used: JMP from SAS

* Split/Partition Graph

A picture containing text

Description automatically generated **Fig**. A.1

* Table

  Description automatically generatedLeaf report highlighting the leaf rules and probabilities at each node, **Fig**. A.2 below
* **Table

  Description automatically generated**Confusion matrix and column contributions
* Overall misclassification rate = 11%
* Misclassification rate for ‘developed’ = 28%
* Misclassification rate for ‘developing’ = 7%
* No. of splits = 10
* No. of terminal nodes = 11

**Graphical user interface, text, application

Description automatically generated Fig**. A.3

* Classification Tree diagram showing 10 split levels, conditions for each split based on below 3 variables: **Fig** A.4 below
  + GDP per capita
  + Population density
  + % Population having access to electricity

Timeline

Description automatically generated

1. **Cluster analysis for agricultural analysis of country (K-means clustering)**

Software used: Minitab and JMP from SAS

Input variables used:

* Agricultural land (in sq. km.)
* Agricultural imports
* Agricultural exports

Table

Description automatically generatedTable

Description automatically generatedDeveloped countries, 2015 (**Fig**. A.5) Developed countries, 1975(**Fig**. A.6)

Table

Description automatically generated

Table

Description automatically generatedTable

Description automatically generatedTable

Description automatically generated

**Table

Description automatically generated**Chart, bar chart, histogram

Description automatically generatedDendrogram: developed countries, 2015, countries forming unique clusters: **Fig.** A.7 and A.8

Graphical user interface, application, table, Excel

Description automatically generated**Chart, histogram

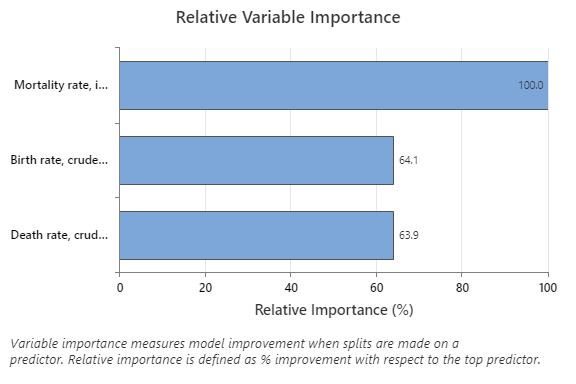
Description automatically generated**Dendrogram: developed countries, 1975, countries forming unique clusters: **Fig.** A.9 and A.10

1. **Regression tree for prediction of life expectancy trends of country**

Software used: Minitab and JMP from SAS

* Regression Tree diagram below, conditions for each split based on below 3 variables:
  + Birth rate (per 1000 people)
  + Death rate (per 1000 people)
  + Mortality rate (per 1000 people)

Below are the relevant regression tree outputs for developed country life expectancy data.

**Table

Description automatically generated For developed countries: Fig.** A.11

**Chart, line chart

Description automatically generated**

**Table

Description automatically generated with medium confidenceTable 1- Leaf report –** prediction of life expectancy for developed countries

|  |  |
| --- | --- |
| **Leaf rule** | **Mean life expectancy years** |
| Birth rate <= 15.3515, Death rate <= 3.802, 8.55 < Mortality rate <= 49.9 | 42.44 |
| Birth rate <= 13.2695, Death rate <= 4.4495, 4.8 < Mortality rate <= 8.55 | 48.96 |
| 82.7 < Mortality rate <= 83.9 | 50.00 |
| Mortality rate > 83.9 | 59.41 |
| Birth rate <= 13.2695, Death rate <= 4.4495, Mortality rate <= 4.8 | 59.46 |
| Birth rate > 30.605, 49.9 < Mortality rate <= 82.7 | 62.39 |
| Birth rate <= 30.605, 49.9 < Mortality rate <= 82.7 | 66.23 |
| Death rate > 3.802, 20.65 < Mortality rate <= 49.9 | 69.58 |
| Death rate > 3.802, 13.85 < Mortality rate <= 20.65 | 71.54 |
| Birth rate > 15.3515, Death rate <= 3.802, 8.55 < Mortality rate <= 49.9 | 73.24 |
| Death rate > 3.802, 8.55 < Mortality rate <= 13.85 | 73.90 |
| Death rate > 11.15, 4.45 < Mortality rate <= 8.55 | 74.73 |
| Birth rate > 13.2695, Death rate <= 4.4495, Mortality rate <= 8.55 | 76.47 |
| 4.4495 < Death rate <= 11.15, 4.45 < Mortality rate <= 8.55 | 77.42 |
| Death rate > 4.4495, Mortality rate <= 4.45 | 80.35 |

Diagram, engineering drawing

Description automatically generatedRegression tree showing 15 terminal nodes: ‘developed’ countries prediction of life expectancy, **Fig.** A.12

**For developing countries: Fig.** A.13

**Chart, bar chart

Description automatically generatedTable

Description automatically generated**

**Table

Description automatically generatedChart

Description automatically generated**

**Table 2- Leaf report –** prediction of life expectancy for developing countries

|  |  |
| --- | --- |
| **Leaf rule** | **Mean life expectancy years** |
| Death rate > 37.4215, Mortality rate > 64.85 | 20.32 |
| 23.978 < Death rate <= 37.4215, 64.85 < Mortality rate <= 226 | 39.10 |
| 20.7415 < Death rate <= 23.978, 64.85 < Mortality rate <= 226 | 43.72 |
| 18.275 < Death rate <= 20.7415, Mortality rate > 64.85 | 46.55 |
| 16.0365 < Death rate <= 18.275, Mortality rate > 64.85 | 49.86 |
| 14.035 < Death rate <= 16.0365, Mortality rate > 64.85 | 52.56 |
| 11.9695 < Death rate <= 14.035, Mortality rate > 64.85 | 55.28 |
| Death rate > 10.123, 34.25 < Mortality rate <= 64.85 | 57.96 |
| 10.176 < Death rate <= 11.9695, Mortality rate > 64.85 | 58.26 |
| 20.7415 < Death rate <= 37.4215, Mortality rate > 226 | 58.81 |
| Death rate <= 10.176, Mortality rate > 64.85 | 62.21 |
| 8.307 < Death rate <= 10.123, 34.25 < Mortality rate <= 64.85 | 62.41 |
| 6.703 < Death rate <= 8.307, 34.25 < Mortality rate <= 64.85 | 65.24 |
| Death rate <= 6.703, 34.25 < Mortality rate <= 64.85 | 68.10 |
| 23.85 < Mortality rate <= 34.25 | 69.25 |
| 15.65 < Mortality rate <= 23.85 | 71.30 |
| 10.05 < Mortality rate <= 15.65 | 72.92 |
| Mortality rate <= 10.05 | 76.06 |

Regression tree showing 18 terminal nodes**Diagram

Description automatically generated:** ‘developing’ countries prediction of life expectancy, **Fig.** A.14