Data Science, Machine Learning and Statistics

Instructions: Click on the link to access each author's presentation.

Chair: Manuel Cuéllar

Participants:

Gonzalo Pérez de la Cruz: Classification using probabilistic models and other methods for mixed data

Haile Mekonnen Fenta: Particulate Matter drives Acute Respiratory Infection among Under-Five Children across sub-Saharan African Countries: Machine Learning Approaches

Miguel David Álvarez Hernández:* Innovations and Strategies in the Mexican Electoral Registry through Data Science

Issoufou Seidou Sanda: Enhancing the Quality of Online Self-Paced Training on the SEEA Using Large Language Models and Interactive Knowledge Graphs

*Work presentation not available or non-existent

IAOS-ISI 2024

MEXICO CONFERENCE

CLASSIFICATION USING PROBABILISTIC MODELS AND OTHER METHODS FOR MIXED DATA

INEGI

International Statistical Institute

IAOS

IMPROVING OFFICIAL STATISTICS

Classification using probabilistic models and other methods for mixed data

Gonzalo Pérez and Guillermina Eslava

IOS-ISI 2024
MEXICO CONFERENCE

Abstract

In this work we present comparative results of some classification methods for the case of two populations and a set of binary and continuous variables.

The goal is to analyze whether using discriminant analysis based on probabilistic graphical models performs well when **the sample size is small**.

The performance of the methods is compared in terms of classification error rates on both simulated and real data.

Notation

Π_c: Population $C \in \{1, 2\}$.

$x = (x_1, \ldots, x_p) = (i_1, ..., i_q, y_1, ..., y_r)$: mixture of $q + r = p$ random variables.

Set of labelled independent observations $\{C_i, x_i\}, i \in \{1, \ldots, n\}$.

Probability that an observed individual belongs to class Π_c:

$$\pi_1 = P(C = 1),\ \pi_2 = P(C = 2)$$

Probability that an individual belongs to class Π_c given the observed value of x:

$$P(C = 1|x),\ P(C = 2|x)$$

IOS-ISI 2024 MEXICO CONFERENCE

Classification rule

Assign an observation x to the class with highest probability:

$$x \mapsto \quad argmax_{c\in\{1,2\}} P(C = c|x) \tag{1}$$

If one assumes that x has an *arbitrary* density $f_c(x|C = c)$ in population $C : 1, 2$, (1) is equivalent to

$$x \mapsto \Pi_1 \quad \text{if} \quad \log \frac{f_1(x|C = 1)}{f_2(x|C = 2)} > \log \frac{\pi_2}{\pi_1}. \tag{2}$$

Probabilistic graphical models

We can use probabilistic graphical models for $f_c(x|C=c)$: Markov networks or Bayesian networks.

Mixed graphical models (MGM), Lauritzen and Wermuth (1989), are Markov Networks used to model mixtures of variables and they are based on the conditional Gaussian distribution.

Consider a set of p variables $V = \Delta \cup \Gamma$, where $\Delta = \{i_1, \ldots, i_q\}$ is a set of q binary variables and $\Gamma = \{y_1, \ldots, y_r\}$ a set of r continuous variables, $x = (i, y)$.

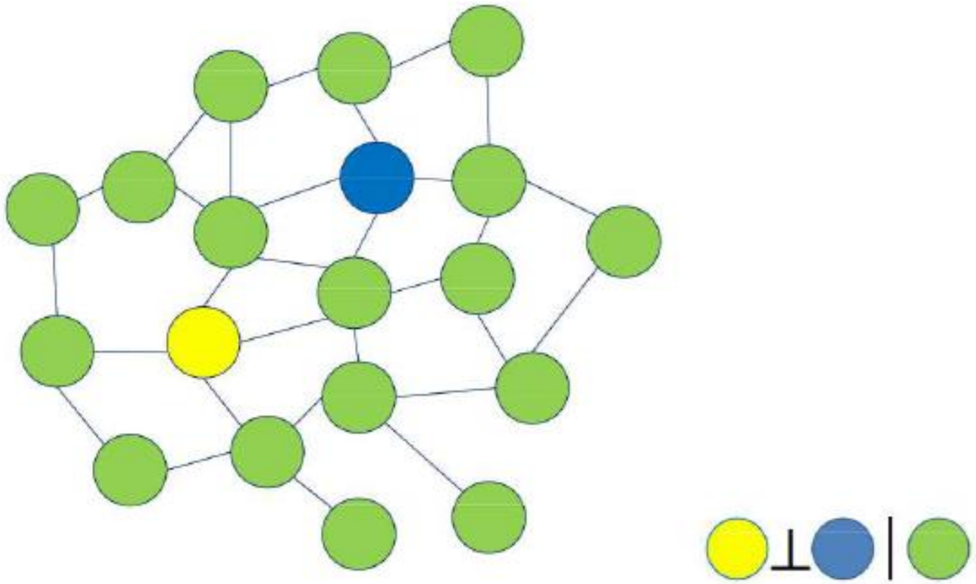
$$f(x) = f(i, y) = p(i) f(y|i), \tag{3}$$

where $p(i) = P(x_\Delta = i) > 0$ corresponds to a positive multinomial distribution and $f(y|i)$ to a Gaussian density $N(\mu(i), \Sigma(i))$ for each $i \in \mathcal{I}$, $|\mathcal{I}| = 2^q$.

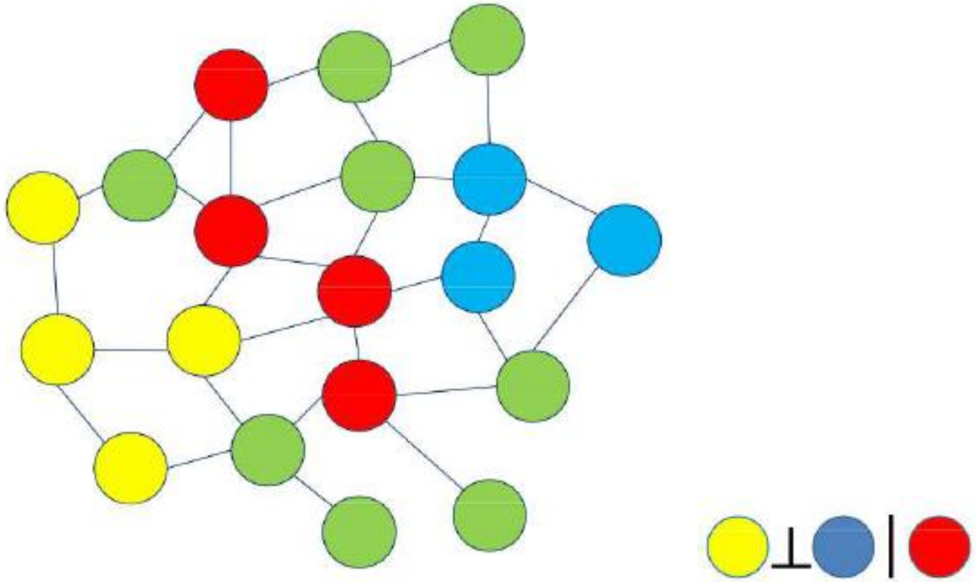
Probabilistic graphical models

In these models $f(x)$ satisfies the Markov properties with respect to an undirected marked graph $G = (V, E)$.

$$(i, j) \notin E \iff x_i \perp\!\!\!\perp x_j | \{x_1, \ldots, x_p\} \setminus \{x_i, x_j\}.$$



Pairwise Markov property



Global Markov property

Peter Green Conference at the ISI World Statistics Congress, Dublin, 2011.

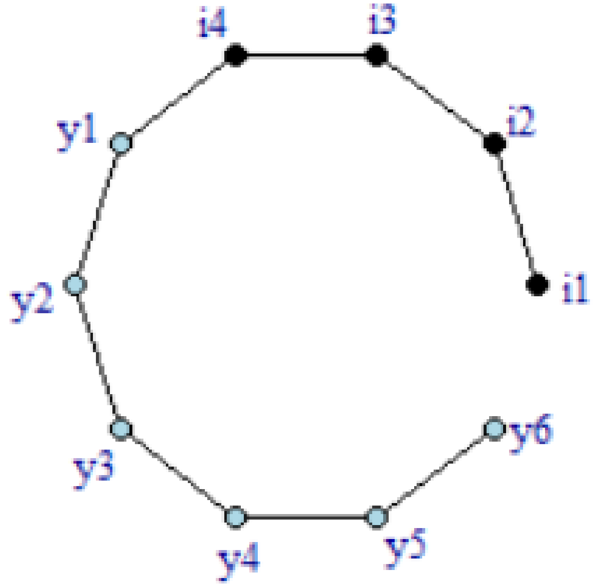
Estimation:

$$f_{\widehat{G}} \rightarrow \widehat{f}_{\widehat{G}}$$

A graph is *decomposable* iff it is triangulated and does not contain any path between two non-adjacent discrete vertices passing through only continuous vertices (strong restriction).

Decomposable tree models

A tree is a (connected) graph without cycles.



Estimation:

$\tau_c \in T$ is estimated/identified/learned/selected using *minForest*{*gRapHD*}

The maximum likelihood estimator of the density also factorizes as

$$
\begin{aligned}
\widehat{f}(i,y) &= \widehat{p}(2\pi)^{-\frac{r}{2}} \mid \widehat{\Sigma}(i)) \mid^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(y-\widehat{\mu}(i))^t \widehat{\Sigma}(i)^{-1}(y-\widehat{\mu}(i))\right\} \\
&= \prod_{j=1}^{k} \frac{\widehat{f}_{[C_j]}(x_{C_j})}{\widehat{f}_{[S_j]}(x_{S_j})} \qquad (6) \\
&= \prod_{j\in V} \widehat{f}(x_j \mid x_{pa_j}). \qquad (7)
\end{aligned}
$$

This factorization allows to estimate $f(i,y)$ through its estimated factors $\widehat{f}(x_j \mid x_{pa_j})$ without calculating $\left(\widehat{p}(i), \widehat{\mu}(i), \widehat{\Sigma}(i)\right)$.

Estimated Classification rule

We use the Bayes rule (2) with two *tree-structured decomposable* CG densities, $f_{\mathcal{T}_1}(i, y)$ and $f_{\mathcal{T}_2}(i, y)$, as classification rule, i.e.

$$x = (i, y) \mapsto \Pi_1 \quad \text{if} \quad \log \frac{f_{\mathcal{T}_1}(i, y)}{f_{\mathcal{T}_2}(i, y)} > \log \frac{\pi_2}{\pi_1}, \tag{8}$$

with ML estimated densities:

$$\log \frac{\widehat{f}_{\mathcal{T}_1}(i, y)}{\widehat{f}_{\mathcal{T}_2}(i, y)} > \log \frac{\widehat{\pi}_2}{\widehat{\pi}_1}. \tag{9}$$

$\mathcal{T}_c \in T$ is estimated/identified/learned/selected using *minForest*$\{$*gRapHD*$\}$
$\widehat{f}_{\mathcal{T}_c}(i, y)$ is computed with *bn.fit*$\{$*bnlearn*$\}$
$\widehat{\pi}_1$ and $\widehat{\pi}_2$ are the relative sample sizes of each population.

Empirical performance

Classification Method	Name	Classification Method	Name
Methods with no interactions		*Methods with pairwise interactions*	
Linear discriminant analysis	LDA	Tree-structured discriminant	CG-tree
Penalized LDA	LDA-pen	Forest-structured	CG-forest
Naive	Naive	Step reduced Tree-structured .	CG-tree-step
Logistic regression	LOG	Step reduced Forest-structured	CG-forest -step
Logistic lasso	LOG-lasso	LDA with pairwise int.	LDA2
SVM with linear kernel	SVM-lin	Step reduced Logistic with pairwise int.	LOG2-step
		Logistic lasso with pairwise int.	LOG2-lasso
		Quadratic discriminant analysis	QDA
		SVM with polynomial kernel	SVM-poly2
Algorithmic methods		*Deep neural networks*	
K nearest neighbour	k-nn	Deep neural networks (`Keras`)	DNN
Random forests	Rand-forest	DNN with variable selection (`PyTorch`)	LassoNet

Table 1: Classification methods used to compare the performance of the tree-structure discriminant analysis.

Figure 1: *Breast cancer dataset* with $n_{success} = 99$ and $n_{failure} = 87$. Estimated test and training error rates. Values averaged across 1000 random training-test data splits, except for DDN and LassoNet with 50 and 100 data splits, respectively. The data splits were done within each group in proportions $(9/10, 1/10)$.

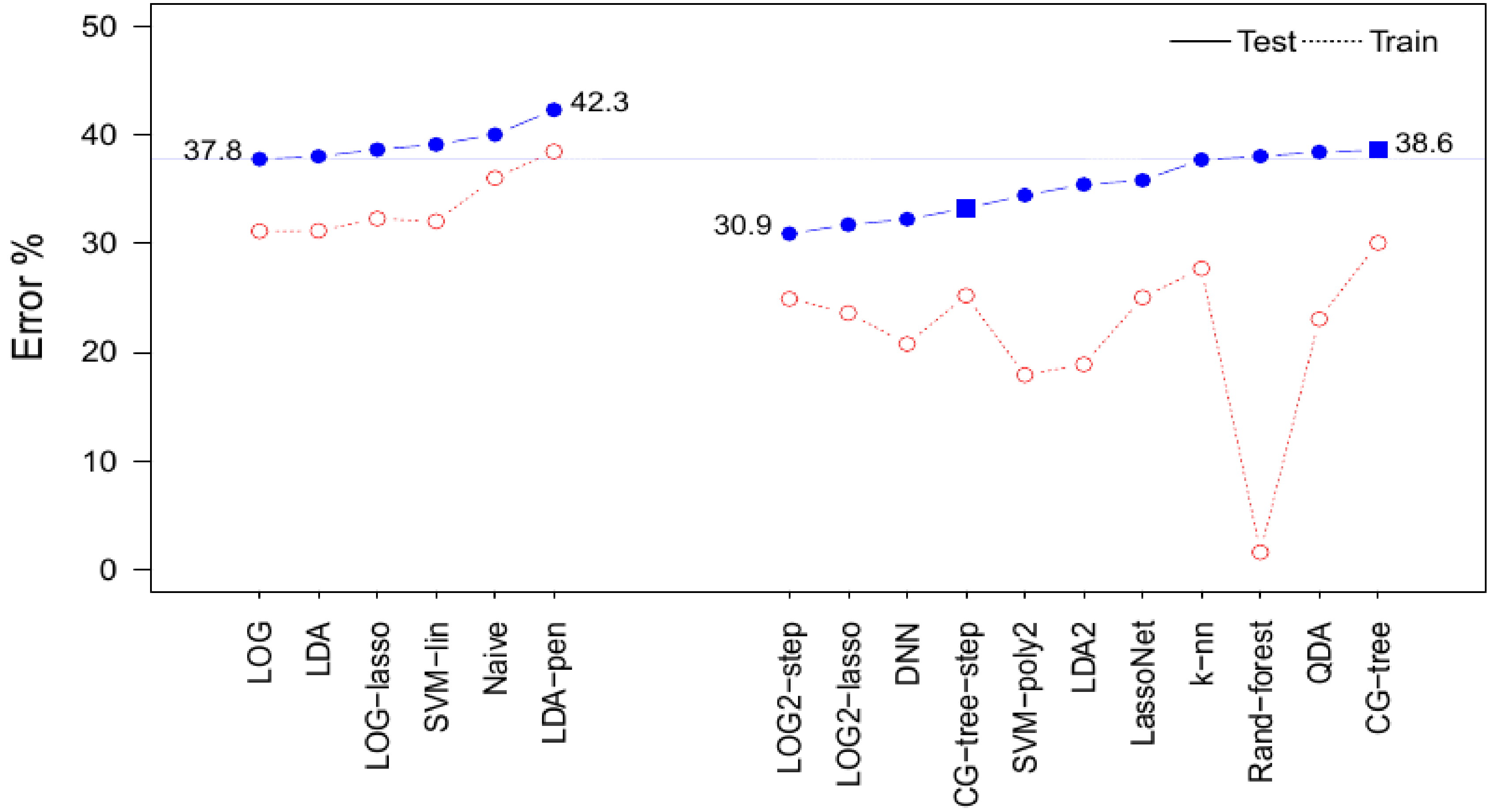
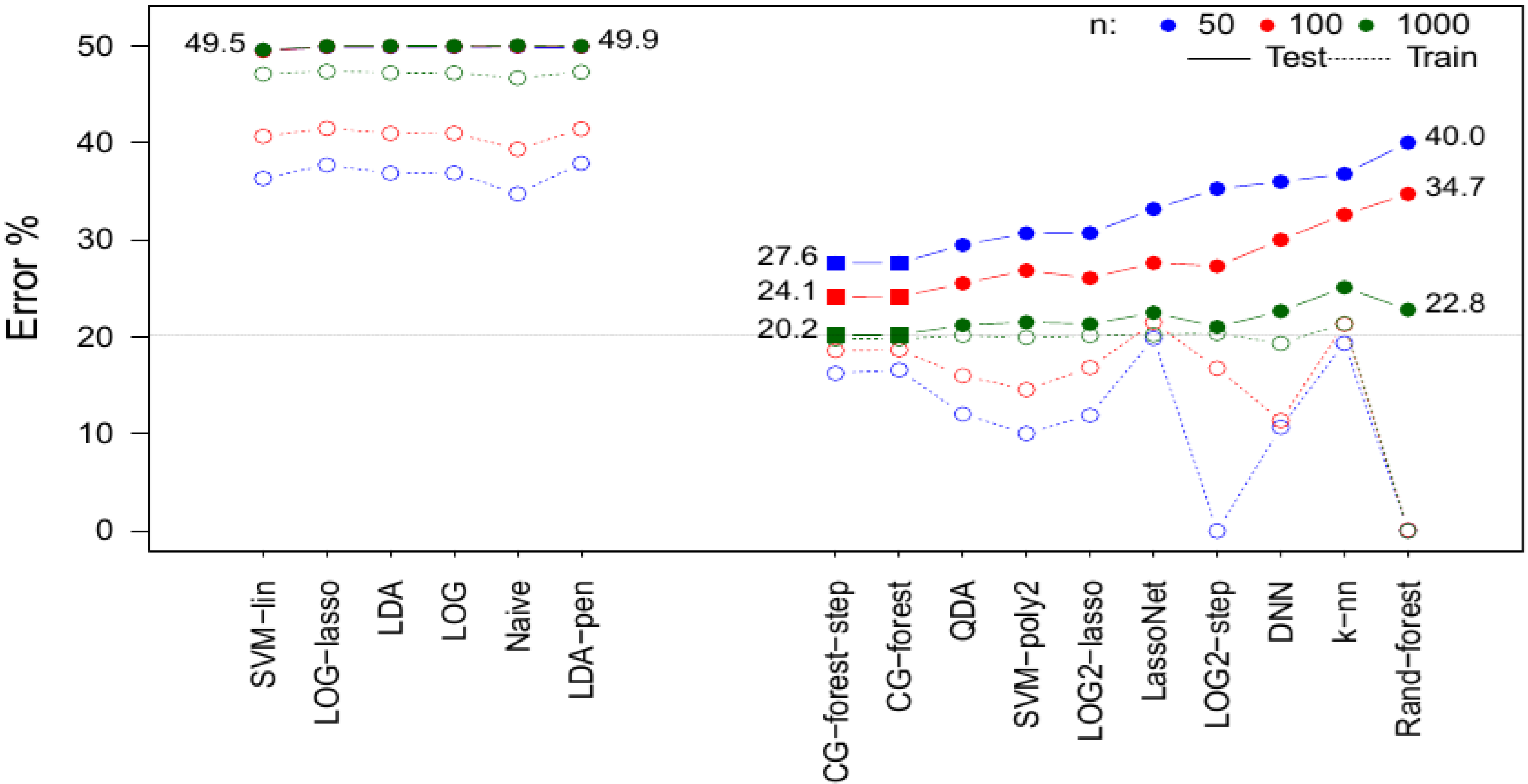


Figure 2: Simulated data. Estimated test and training error rates. Test errors averaged across 1000 test sets of size 1000, except for DDN and LassoNet with 50 and 100 test sets, respectively.

$E_1(x_j) = E_2(x_j)$, $V_1(x_j) = V_2(x_j)$, $j = 1, \ldots, 10$; $Cor_1(x_i, x_{i+1}) > 0$ and $Cor_2(x_i, x_{i+1}) < 0$, $\forall\ i \in \{1, \ldots, 9\}$



Comments

Deep neural networks did not perform the best in any of the data sets. For high-dimensional small tabular data sets, it has been noticed that they do not perform well, Margeloiu et al. (2023).

The proposed method with variable selection and logistic regression with interactions and lasso had a good performance in the simulation and the real data sets. They are straightforward to apply and not highly computer-intensive. This makes them worth considering for classification, especially for small sample sizes where parsimonious methods with variable selection or regularization might perform better.

References

Mixed Graphical Models

Lauritzen, S.L. (1996). Graphical Models. Clarendon Press, Oxford

Højsgaard, S., Lauritzen, S.L., Edwards, D. (2012). Graphical Models with R. Springer, New York

Scutari, M. (2021). Bayesian Networks - With examples in R. 2nd ed.

Mixed Graphical Models and Classification

Krzanowski, W.J. (1975). Discrimination and classification using both binary and continuous variables. Journal of the American Statistical Association, 70(352), 782–790.

Krzanowski, W.J. (1994). Quadratic location discriminant functions for mixed categorical and continuous data. Statistics & Probability Letters,19(2), 91–95

Thank you

Particulate Matter drives Acute Respiratory Infection among Under-Five Children Across sub-Saharan African Countries: Machine Learning Approaches for Large Datasets

Haile Mekonnen[1,4], Temesgen Z[2] , Saloshni Naidoo[3], Rajen N[4]. Naidoo, and Henry Mwambi[5]

[1]Discipline of Public Health Medicine, School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[2]School of Mathematics, Statistics and Computer Science, College of Agriculture Engineering and Science, University of KwaZulu-Natal, Durban, South Africa

[3]Discipline of Occupational and Environmental Health, School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[4]Department of Statistics, College of Science, Bahir Dar University, Bahir Dar- Ethiopia

Contentes

Introduction

- ✓ Acute Respiratory Infections are common illnesses and the leading cause of death among children under the age of five.
- ✓ Under-five children are at greater risk than the other population groups from many of the adverse health effects of air pollution
- ✓ Under-five death rate due to ARIs was 73/1000 in Africa and and 9/1000- in Asia live births (LMICs)
- ✓ More than 89% of deaths due to air pollution occurred in LMICs, mainly in Africa and Asia.
- ✓ Africa accounts for the highest excess mortality from ambient air pollution among under-five children, to which ARIs were suggested as a potential contributor
- ✓ No studies use ML algorithms to investigate the effects of air pollutants and climate factors on ARIs on big data context.
- ✓ We used ML techniques to select and identify the associated risk factors with symptoms of ARIs in sSA countries.

Data sources and variables

- ✓ The data for this study came from two sources: DHS (33 sSA countries) and the NASA: PM 2.5 concentration and the nitrogen dioxide (NO_2) was estimated form of raster images (GeoTIFF).

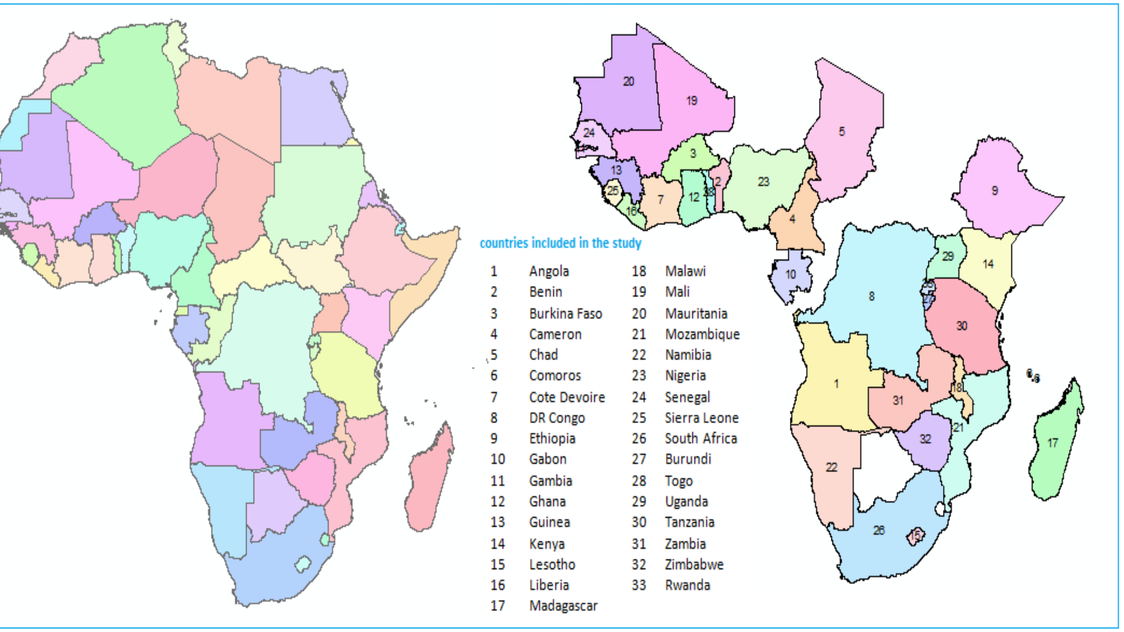


Figure 1: Eligible sub-Saharan African countries include

Table 1: Selection of study participants from 33 sSA countries with recent DHS reports from 2012 to 2022

A total of 49 countries are located in Sub-Saharan Africa			
East African regions 18 countries	West African regions 17 countries	Central Africa regions 9 countries	Southern Africa regions 5 countries
A total of 16 countries were excluded for the following reasons			
6 countries were excluded. ✓ 3 countries no DHS report ✓ 3 countries no GPS is available	4 countries were excluded. ✓ 1 country with no DHS report ✓ 3 countries no GPS available	4 countries were excluded. ✓ 1 country with no DHS report ✓ 3 countries no GPS available	2 countries were excluded. ✓ 2 countries where no GPS is available
A total of 33 countries included			
East African regions ✓ 12 countries (Burundi, Comoros, Ethiopia, Madagascar, Malawi, Mozambique, Rwanda, Tanzania, Uganda, Zambia, Zimbabwe, and Kenya) ✓ 7,595 PSU ✓ 124,106 U5C	West African regions ✓ 13 countries (Benin, Burkina Faso, Gambia, Ghana, Guinea, Ivery Coast, Liberia, Mali, Mauritania. Nigeria, Senegal, Sierra Leone, and Togo ✓ 6,952 PSU ✓ 130,996 U5C	Central African regions ✓ 5 countries (Angola, Cameroon, Chad, Democratic Republic Congo, and Gabon) ✓ 2,546 PSU ✓ 61,911U5C	Southern African regions ✓ 3 countries (South Africa, Lesotho, and Namibia) ✓ 1,595 PSU ✓ 10,494 U5C
A total of ✓ 33 countries ✓ 18,688 PSU ✓ 327,507 U5C			

Variables

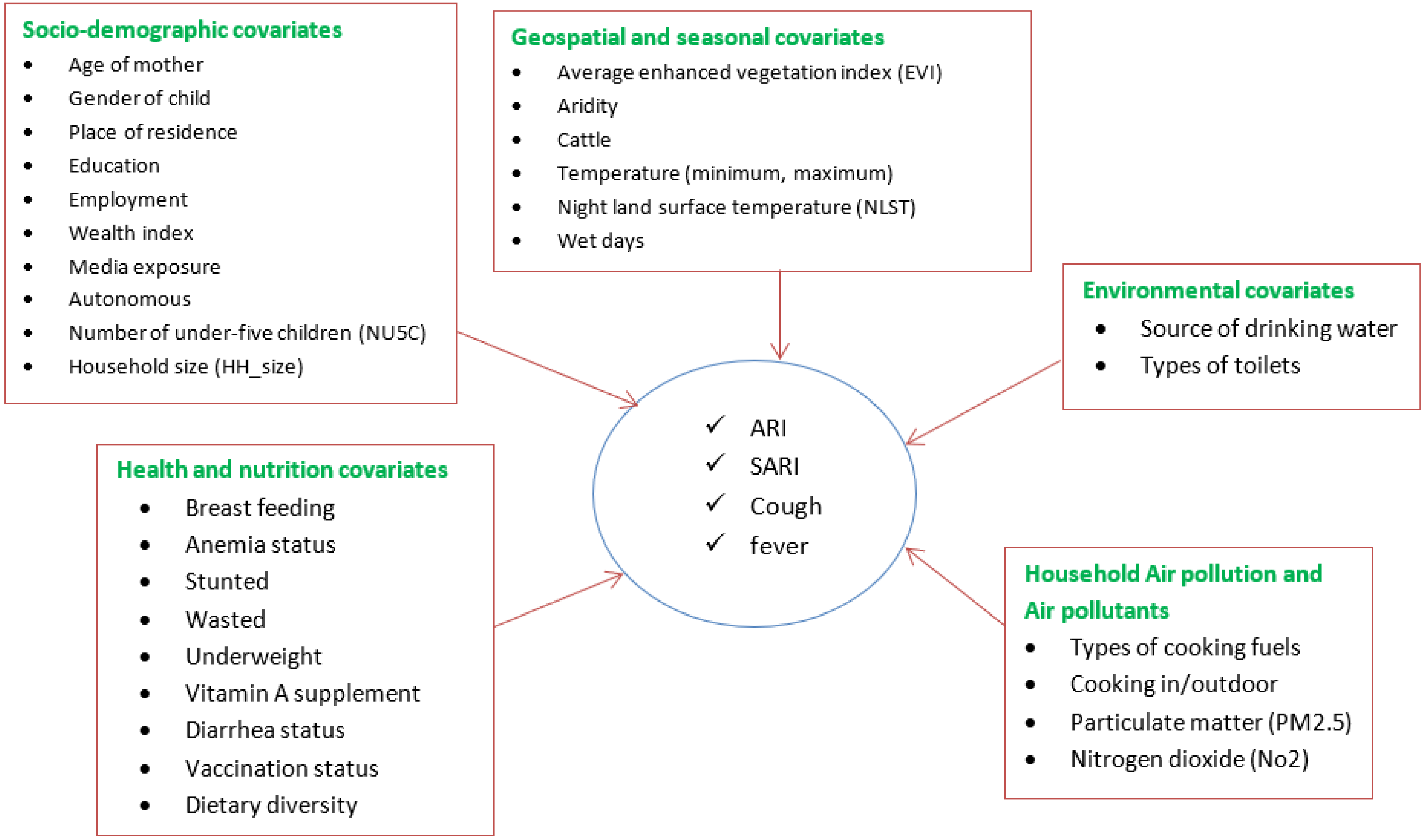
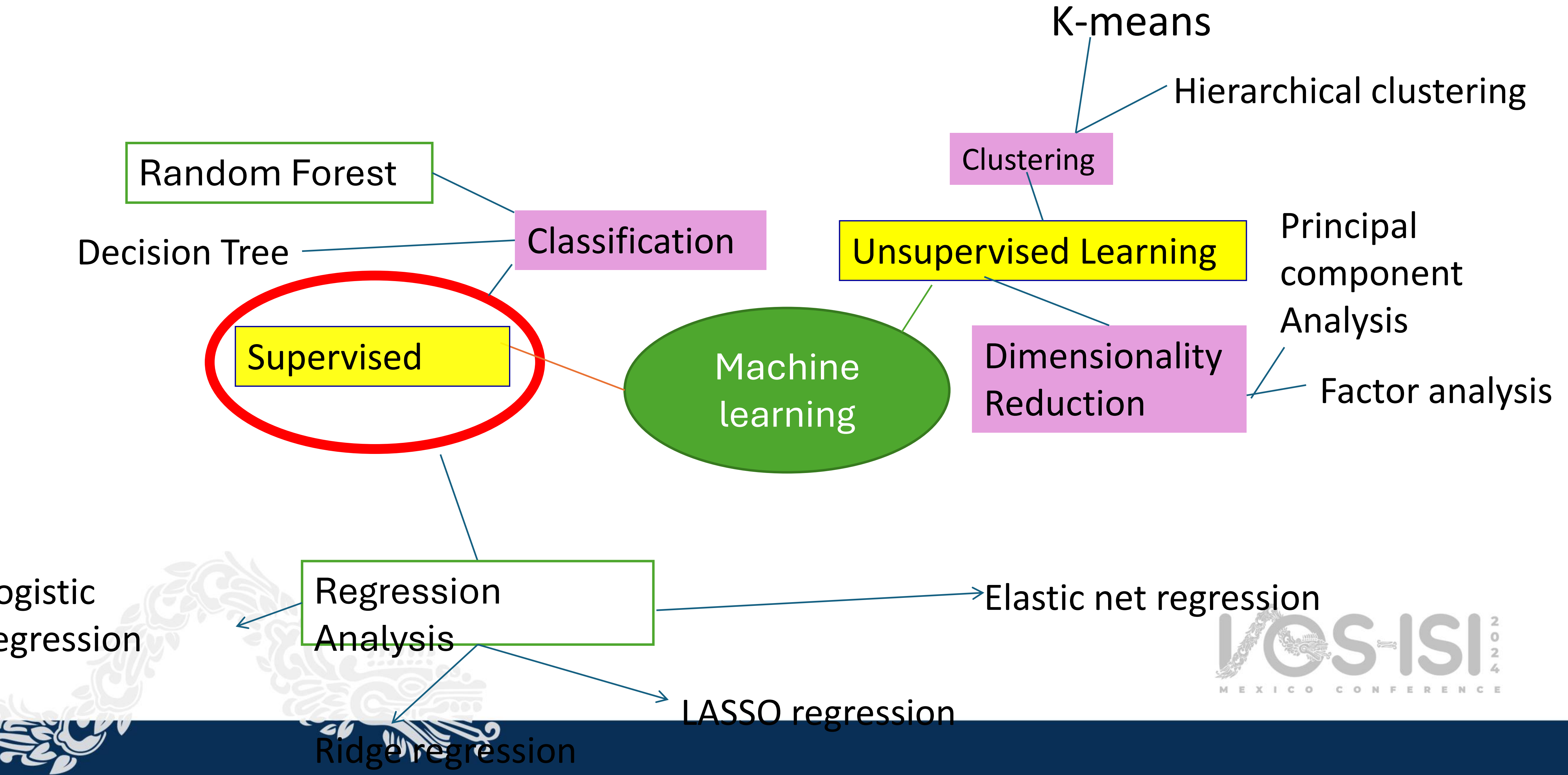


Figure 2: Conceptual framework for features description

Machine Learning Algorithms (MLA)

- Algorithms that learn to make predictions from examples (data)
- used to capture the hidden patterns and factors of outcome variables.
- enable the use of a larger number of predictors, require fewer assumptions, incorporate "multi-dimensional correlations"
- Application: medical sciences for diagnosis and outcome prediction, disease modeling , disease prediction, child mortality predictions, and in industrial applications for prediction and classification.
- Only a few studies on prediction models of childhood for undernutrition .
- Preprocessing: transformation and standardization

Exploratory Data Analysis with Machine Learning Algorithms



Model building

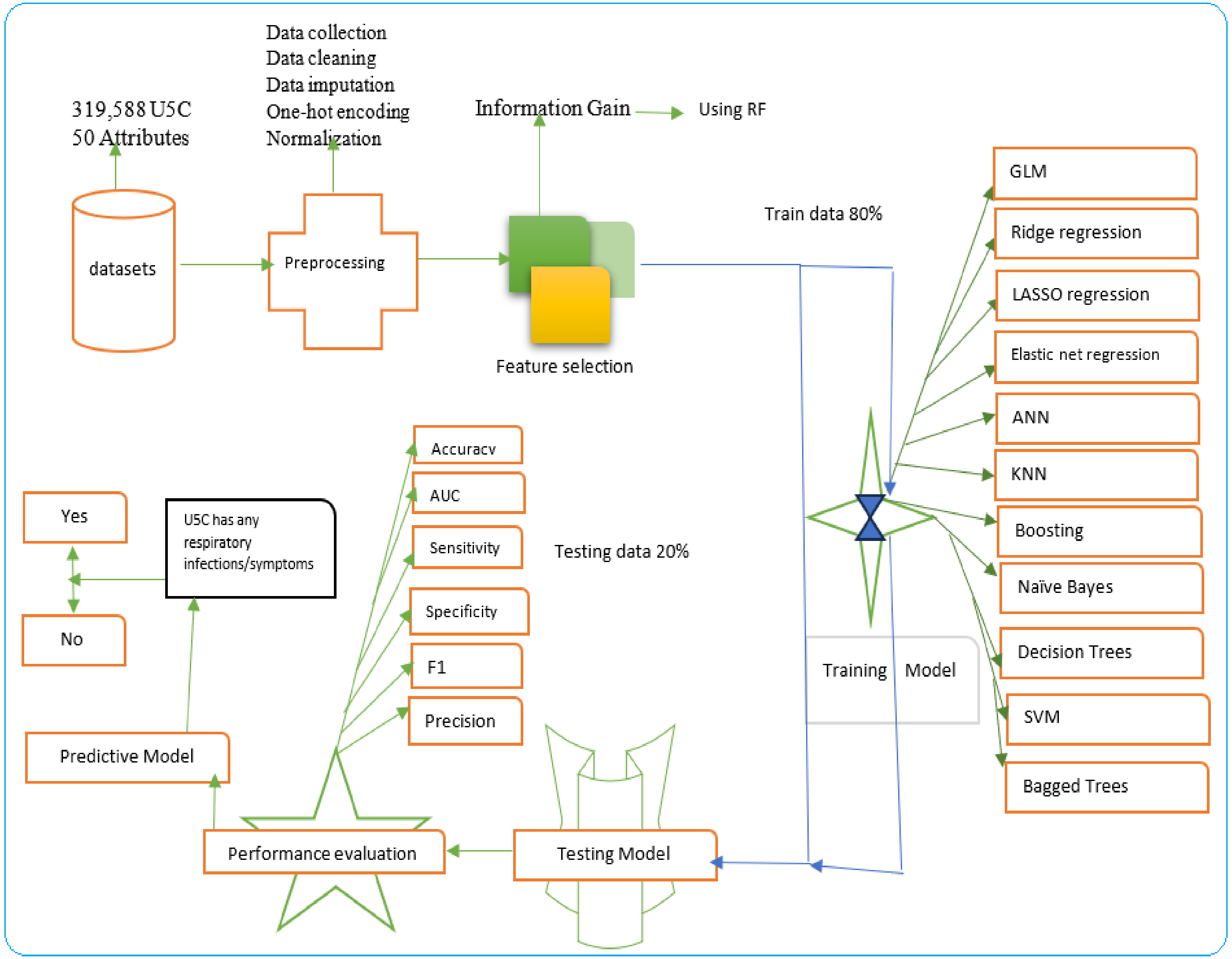


Figure 3: Overview flow chart of the Machine Learning Algorithms used for predicting U5C respiratory infections/symptoms.

Results

The number of under-five children across the DHS waves for each country and the prevalence of symptoms of respiratory infections among U5C children in sSA.

Survey countries	Survey year	Weighted sample	Percent	Children with symptoms of			
				ARI n (%)	SARI n (%)	Cough n (%)	Fever n (%)
Angola	2015	13,439	4.10	606 (4.51)	317 (2.36)	1,416 (10.54)	1,934 (14.39)
Benin	2017	12,529	3.83	702 (5.60)	395 (3.15)	2,016 (16.09)	2,427 (19.37)
Burkina Faso	2021	11,763	3.59	377 (3.20)	230 (1.96)	1,308 (11.12)	2,622 (22.29)
Burundi	2016	12,432	3.80	1,549 (12.46)	1,063 (8.55)	4,740 (38.13)	4,639 (37.31)
Cameroon	2017	8,986	2.74	373 (4.15)	167 (1.86)	1,687 (18.77)	1,387 (15.44)
Chad	2015	16,644	5.08	1,794 (10.78)	1,053 (6.33)	3,092 (18.58)	3,531 (21.21)
Comoros	2011	2,916	0.89	200 (6.86)	130 (4.46)	516 (17.70)	622 (21.33)
Congo democratic	2013	16,960	5.18	2,098 (12.37)	1,244 (7.33)	5,306 (31.29)	5,229 (30.83)
Ivory Coast	2017	9,888	3.02	188 (1.90)	111 (1.12)	1,187 (12.00)	1,724 (17.44)
Ethiopia	2016	9,911	3.03	795 (8.02)	493 (4.97)	1,583 (15.97)	1,354 (13.66)
Gabon	2019	5,882	1.80	233 (3.96)	150 (2.55)	1,426 (24.24)	1,311 (22.29)
Gambia	2019	7,764	2.37	578 (7.44)	288 (3.71)	1,463 (18.84)	1,324 (17.05)
Ghana	2014	5,544	1.69	364 (6.57)	178 (3.21)	744 (13.42)	821 (14.81)
Guinea	2018	6,633	2.03	287 (4.33)	157 (2.37)	744 (11.22)	1,123 (16.93)
Kenya	2022	18,705	5.71	582 (3.11)	340 (1.82)	4,328 (23.14)	3,143 (16.80)
Lesotho	2014	2,818	0.86	259 (9.19)	167 (5.93)	789 (28.00)	405 (14.37)
Liberia	2019	4,083	1.55	518 (10.19)	325 (6.39)	1,379 (27.13)	1,471 (28.94)
Madagascar	2021	11,647	3.56	651 (5.59)	323 (2.77)	2,217 (19.03)	1,438 (12.35)
Malawi	2015	16,209	4.95	1,648 (10.17)	1,044 (6.44)	3,889 (23.99)	4,687 (28.92)
Mali	2018	9,175	2.80	311 (3.39)	189 (2.06)	866 (9.44)	1,497 (16.32)
Mauritania	2019	10,956	3.35	672 (6.13)	495 (4.52)	1,372 (12.52)	1,874 (17.10)
Mozambique	2015	4,954	1.51	758 (15.30)	295 (5.95)	1,415 (28.56)	1,300 (26.24)
Namibia	2013	4,426	1.35	604 (13.65)	380 (8.59)	1,381 (31.20)	1,128 (25.49)
Nigeria	2018	30,597	9.34	1,603 (5.24)	940 (3.07)	4,816 (15.74)	7,535 (24.63)
Rwanda	2019	7,758	2.37	587 (7.57)	351 (4.52)	2,208 (28.46)	1,468 (18.92)
Senegal	2019	5,726	1.75	430 (7.51)	270 (4.72)	848 (14.81)	920 (16.07)
Sierra Leone	2019	8,878	2.71	354 (3.99)	233 (2.62)	1,231 (13.87)	1,473 (16.59)
South Africa	2016	3,250	0.99	150 (4.62)	108 (3.32)	820 (25.23)	647 (19.91)
Tanzania	2022	10,197	3.11	221 (2.17)	145 (1.42)	1,197 (11.74)	1,011 (9.91)
Togo	2013	6,460	1.97	922 (14.27)	498 (7.71)	1,698 (26.28)	1,413 (21.87)
Uganda	2016	14,378	4.39	2,164 (15.05)	1,349 (9.38)	5,766 (40.10)	5,027 (34.96)
Zambia	2019	9,308	2.84	241 (2.59)	142 (1.53)	1,948 (20.93)	1,549 (16.64)
Zimbabwe	2015	5,3691	1.74	445 (7.82)	166 (2.92)	2,103 (36.95)	796(13.99)
Total		**327,507**	**100**	23,264 (7.10)	13,736 (4.19)	67,499 (20.61)	68,830 (21.02)

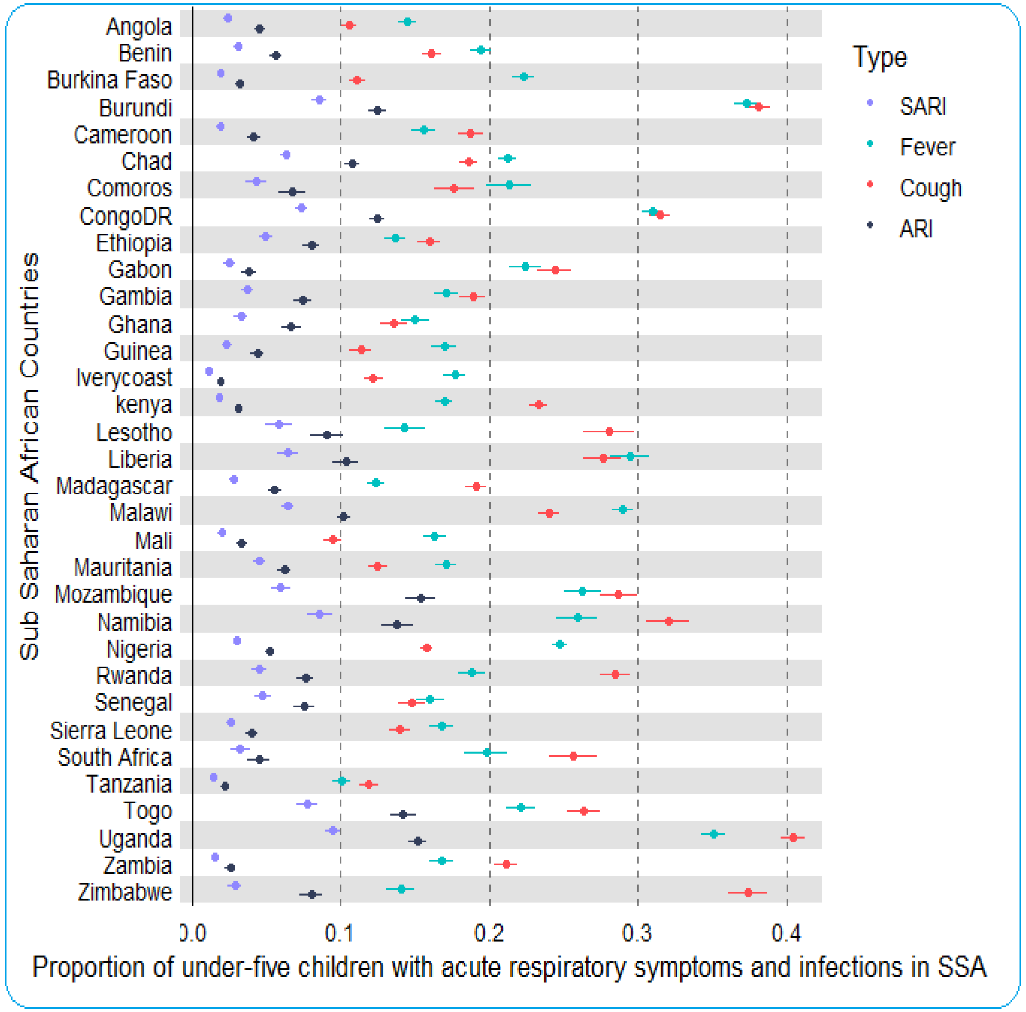
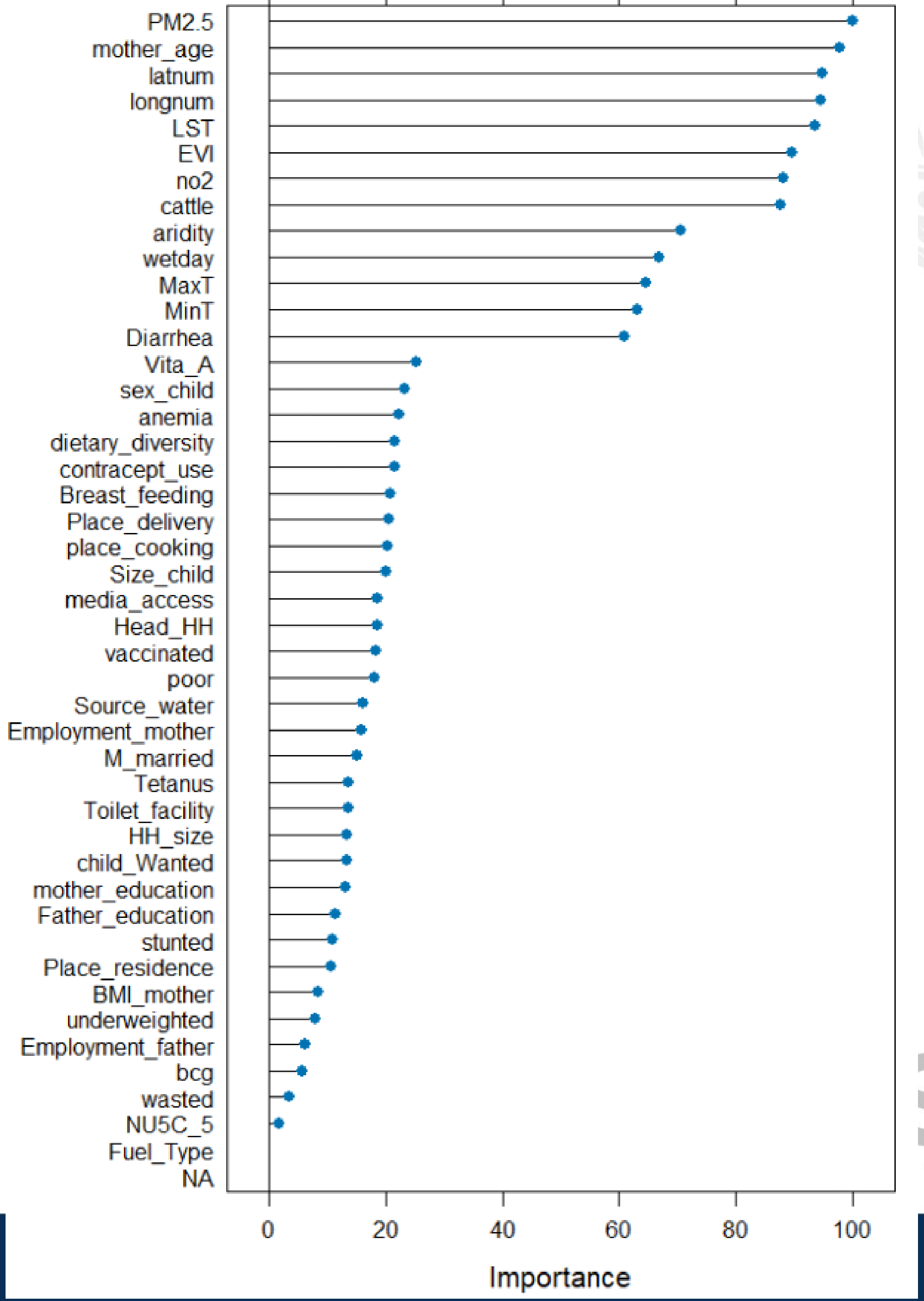


Figure 4: Proportion of under-five children with different AR infections and symptoms across sSA countries

- ✓ The Gini Importance was conducted RF MLA to identify the features on ARIs
- ✓ Features score larger than the second quartile (20.3) was considered as a cut of point for selecting
- ✓ 21 features are retained for the subsequent analysis.
- ✓ Air pollutants and climatic factors: household air pollution and air pollutants have a relative importance score greater than the second quartile (20.3%).
- ✓ Only the mother's age and sex of a child from socio-demographic and diarrhea status and vitamin were selected
- ✓ ML models such as GLM (logistic regression), Ridge, LASSO, Elastic net, ANN, KNN, Boosting, Naïve Bayes, DT, RF, and Bagged Trees were employed



The model evaluation and accuracy

- No substantial difference in accuracies of the different MLAs that can predict the SARIs among U5Cin sSA countries.
- The highest model performance was obtained by Random Forest, Boosting, ANN, and Bagged trees with AUCs of 0.77, 0.76, 0.74, and 0.74 respectively.
- The lowest model performance was observed for DT and NB with AUC=0.68 and 0.70 respectively.
- The RF model is more accurate in distinguishing the diagnosis of SARIs among children under five years

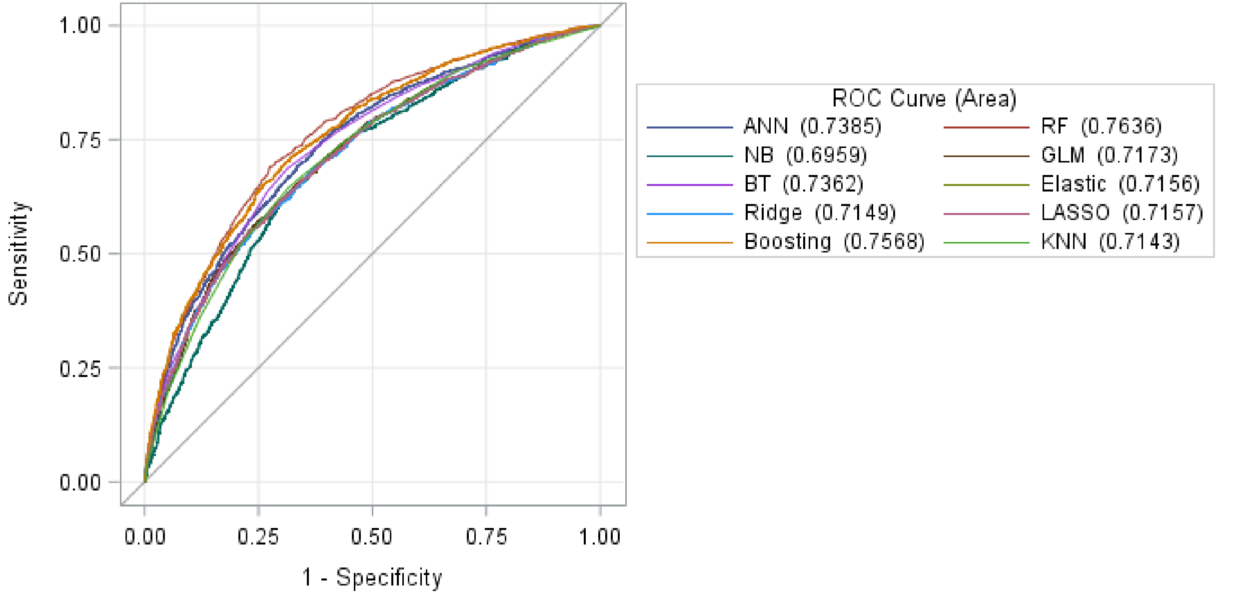


Table 3: The performance of the prediction models based on different classifications using a test dataset with 95% CI

Algorithms	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Accuracy (95 % CI)
GLM	0.64 (0.63, 0.66)	0.67 (0.64, 0.69)	0.72 (0.70, 0.73)	0.65 (0.64, 0.67)
Ridge	0.89 (0.88, 0.90)	0.36 (0.34, 0.39)	0.71 (0.70, 0.73)	0.71 (0.70, 0.73)
Lasso	0.89 (0.88, 0.90)	0.37 (0.34, 0.39)	0.72 (70, 0.73)	0.71 (0.69, 0.72)
elastic-net	0.89 (0.88, 0.90)	0.36 (0.34, 0.39)	0.72 (0.70, 0.73)	0.70 (0.69, 0.72)
ANN	0.64 (0.63, 0.66)	0.71 (0.68, 0.73)	0.74 (0.73, 0.75)	0.67 (0.65, 0.68)
KNN	0.84 (0.83, 0.86)	0.43 (0.40, 0.45)	0.71 (70, 0.73)	0.70 (0.69, 0.72)
NB	0.59 (0.57, 0.61)	0.72 (0.69, 0.74)	0.70 (0.68, 0.71)	0.63 (0.61, 0.65)
Bagged Tree	0.80 (0.78, 0.81)	0.53 (0.50, 0.56)	0.74 (0.72, 0.75)	0.71 (0.69, 0.72)
RF	**0.81 (0.80, 0.83)**	**0.55 (0.52, 0.58)**	**0.77 (0.75, 0.78)**	**0.72 (0.71, 0.73)**
Boosting	0.82 (0.81, 0.84)	0.53 (0.50, 0.55)	0.76 (0.74, 0.77)	0.72 (0.71, 0.74)
DT	0.86 (0.85, 0.88)	0.40 (0.37, 0.73)	0.68 (66, 0.70)	0.71 (0.69, 0.72)

Discussion and Conclusion

- ✓ Dataset contained a total of 51 features and 327,507 under-five children.
- ✓ Parametric linear models, semi-parametric and generalized additive models were used to determine the effects of air pollutants on symptoms of respiratory infections
- ✓ Limited research were used the MLA to determine the association between air pollutants and human health, and none have used ML models to determine the effects of air pollutants on children's symptoms of respiratory infections.
- ✓ The present study attempted to identify the best ML algorithms for the prediction of symptoms of ARI using nationwide cross-sectional data from 33 SSA countries.
- ✓ We started the feature selection process.
- ✓ Using the random forest approach, the ranking of the contributions of the features was determined by using the average Gini Importance method and only 21 features were retained for further ML models.

Discussion and Conclusion

- ✓ Those selected features have scores greater than the second quartile (median), which is used as a rule of thumb for dimension reduction of features.
- ✓ The performances of these ML models were compared using different statistical merits such as sensitivity, specificity, accuracy, and AUC.
- ✓ Climate factors, such as temperature, wet day, and spatial location (longitude, latitude), were among the top features associated with the symptoms of respiratory infections.
- ✓ Spatial location (longitude, latitude) is one of the influential features in predicting and diagnostic symptoms of ARIs
- ✓ PM2.5 was the most influential variable increasing the risk of ARI, together with NO2

Limitations

- ✓ Firstly, we considered only one recent DHS data for each country, and hence we did not model the variables over time.
- ✓ Secondly, the data is cross-sectional so we can only make conclusions on statistical association (not causality).
- ✓ Feature direction: our future focus will be to include the temporal effects to draw inferences over time and possibly causality.

Reference

- ✓ Those selected features have scores greater than the second quartile (median), which is used as a rule of thumb for dimension reduction of features.
- ✓ The performances of these ML models were compared using different statistical merits such as sensitivity, specificity, accuracy, and AUC.
- ✓ Climate factors, such as temperature, wet day, and spatial location (longitude, latitude), were among the top features associated with the symptoms of respiratory infections.
- ✓ Spatial location (longitude, latitude) is one of the influential features in predicting and diagnostic symptoms of ARIs
- ✓ PM2.5 was the most influential variable increasing the risk of ARI, together with NO2

Limitations

- ✓ Firstly, we considered only one recent DHS data for each country, and hence we did not model the variables over time.
- ✓ Secondly, the data is cross-sectional so we can only make conclusions on statistical association (not causality).
- ✓ Feature direction: our future focus will be to include the temporal effects to draw inferences over time and possibly causality.

Thank you

IAOS-ISI 2024 MEXICO CONFERENCE

AI-Assisted Training on System of National Accounts (SNA) and Environmental-Economic Accounting (SEEA)

Enhancing the teaching and practice of SNA and SEEA with advanced technologies

Issoufou Seidou Sanda, Ana Carolina Peixoto Deveza
United Nations Economic Commission for Africa

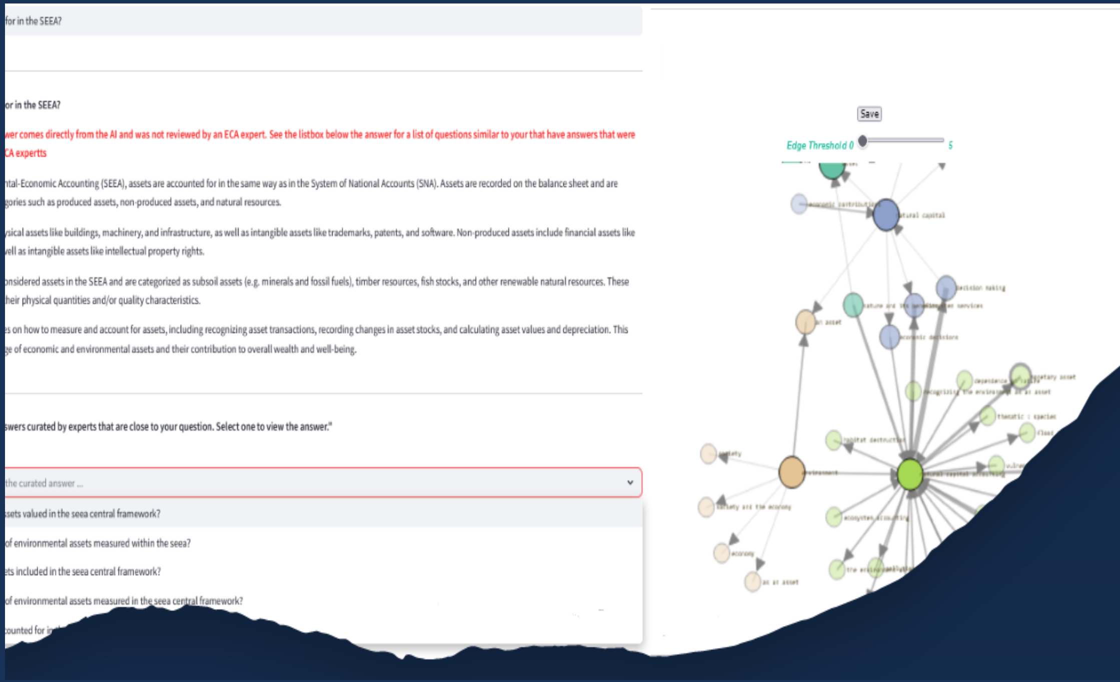
INEGI International Statistical Institute IAOS IMPROVING OFFICIAL STATISTICS

AI-Assisted Training on System of National Accounts (SNA) and Environmental-Economic Accounting (SEEA)

Leveraging Conversational AI for Effective Learning

Plan:

1. The need for harmonized training in Environmental-Economic Accounting (SEEA)
2. How AI can revolutionize learning
3. The potential of knowledge graphs to enhance SEEA learning
4. Perspectives



IOS-ISI 2024
MEXICO CONFERENCE

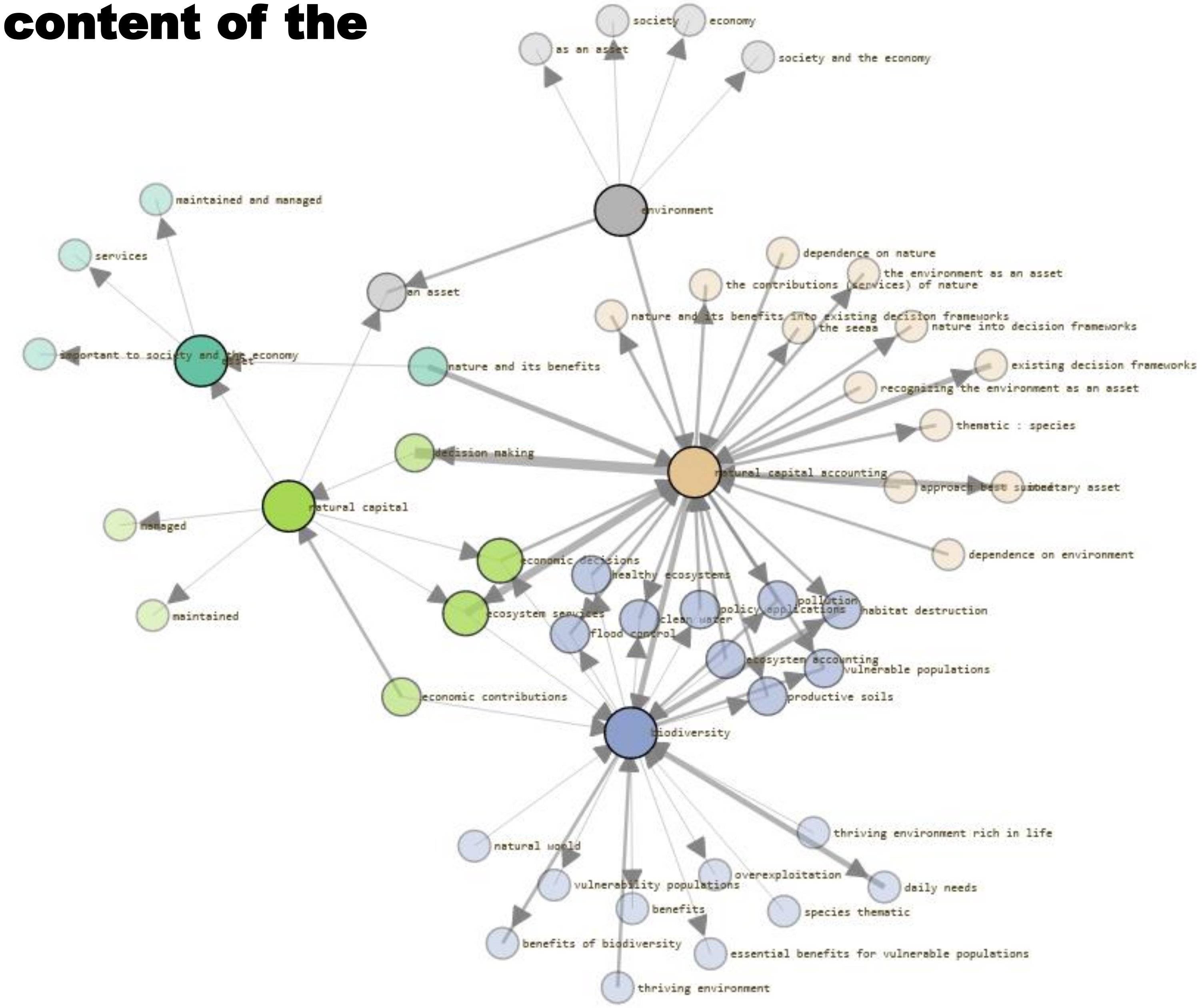
At the heart of the system, there is an AI-built knowledge graph based on the content of the course

The knowledge graph represents the core concepts in the course and how they are related.

Example:

- The course said, “Biodiversity is impacted by habitat destruction.”
- This knowledge is extracted and included in the knowledge graph as a relation between the concept of “biodiversity” and the concept of “habitat destruction”. The nature of the relation is “is impacted by”.
- This relation is represented like this: Biodiversity [- is impacted by - > Habitat destruction
- When a relation is mentioned many times in the course, it means that the relation is important. The relations are, therefore, weighted by the number of times they were mentioned in the course. That allows focusing on the most important relations when there are too many relations.

Because we are using a large language model that understands well human language, it can identify the relation no matter how it was said in the text (the form does not matter, only the meaning matters)



Helping the student navigate in a graph of closely related concepts

AI powered SEEA course assistant - Chapter 4 - Accounting for ecosystem services

Enter your question here:

How are assets measured in the SEEA?

How are assets measured in the SEEA?

Warning: the following answer comes directly from the AI and was not reviewed by an ECA expert. See the listbox below the answer for a list of questions similar to your that have answers that were reviewd and validated by ECA expertts

Assets are measured in the System of Environmental-Economic Accounting (SEEA) using the concept of stocks. Assets are considered as stocks of natural, produced, human, and social capital. These stocks are measured in physical and monetary terms to provide a comprehensive understanding of a country's wealth and well-being. The SEEA also looks at changes in these assets over time to assess sustainability and inform decision-making.

Here are questions with answers curated by experts that are close to your question. Select one to view the answer."

Select the question to see the curated answer ...

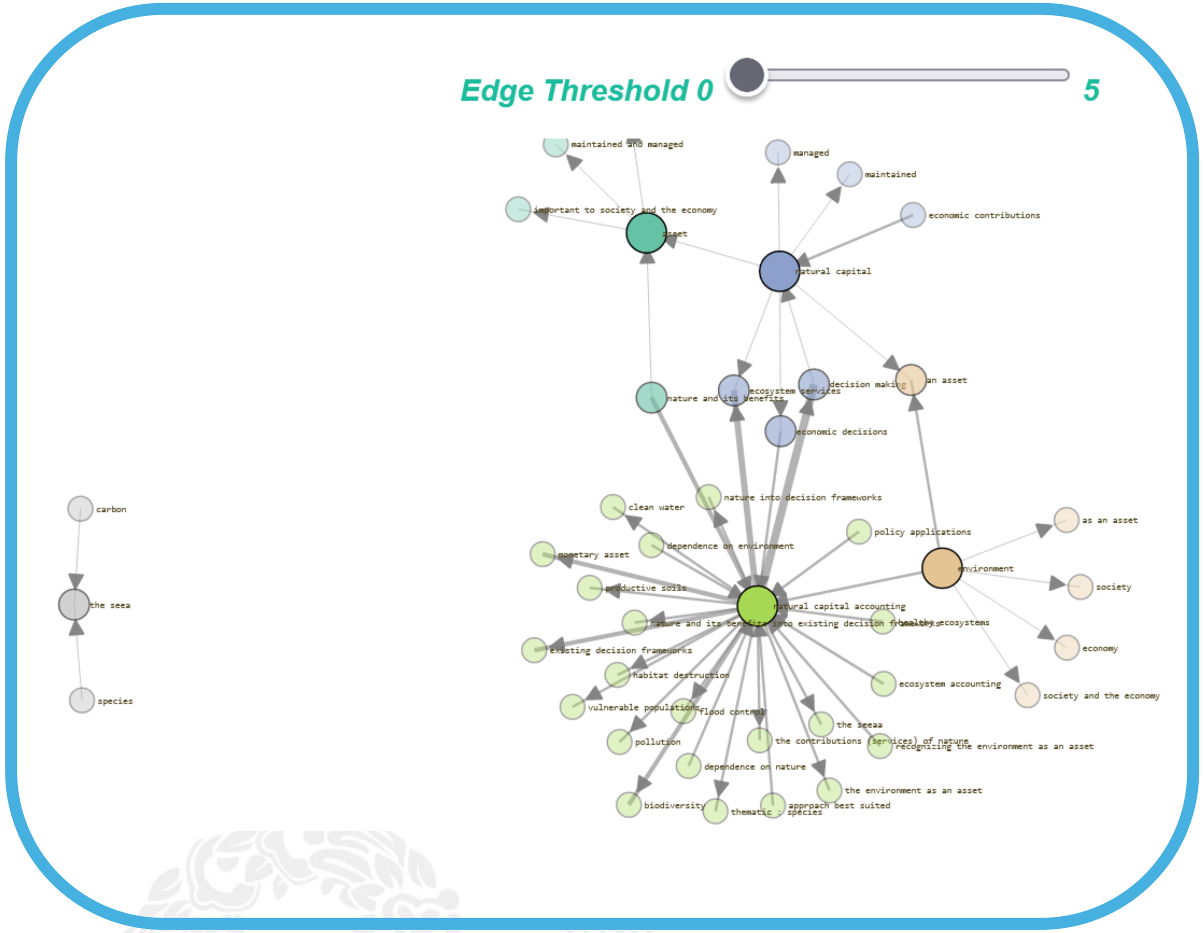
Graphs of concepts that you may want to explore deeper based on your question:

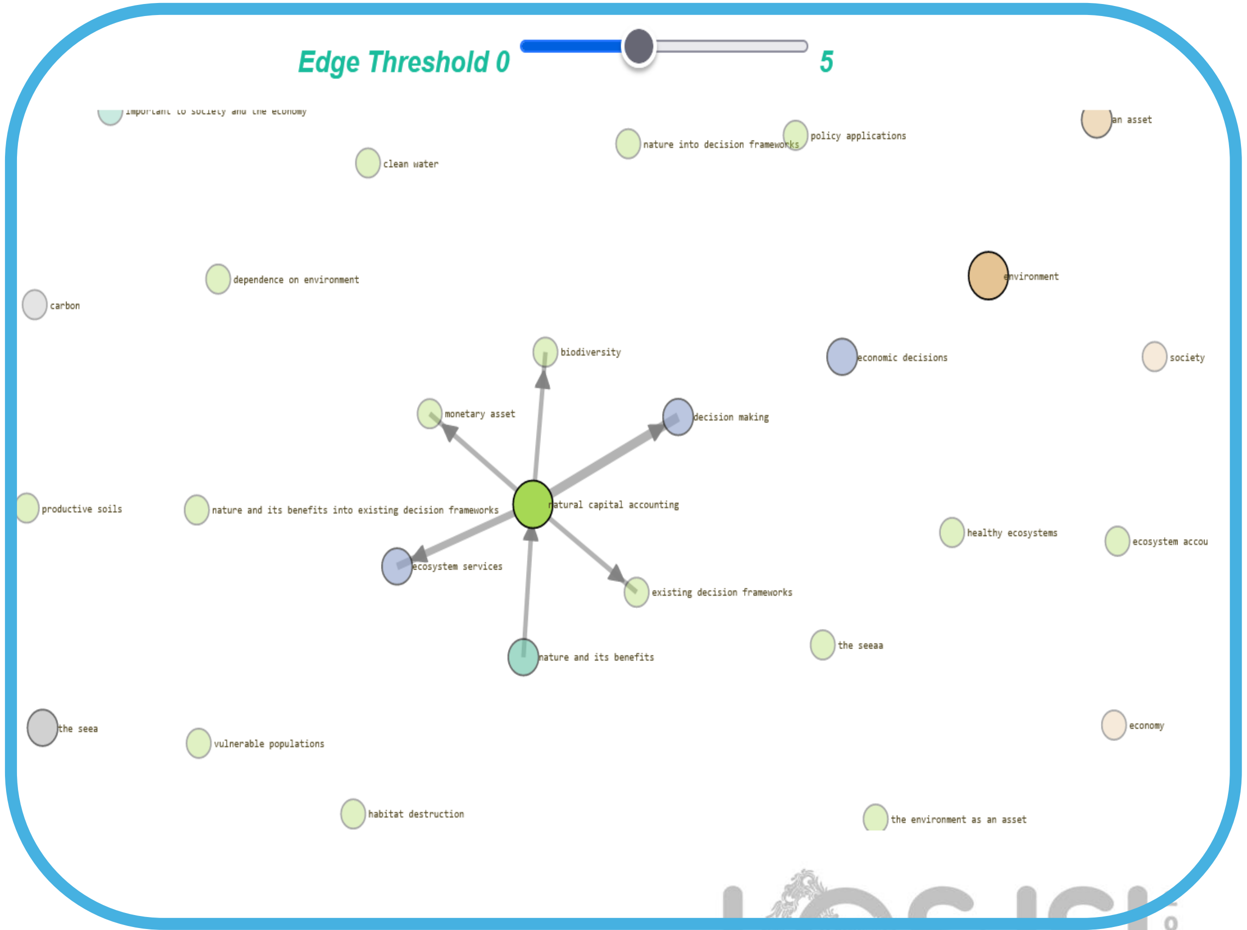
Save

Edge Threshold 0 5

Helping the student navigate in a graph of closely related concepts

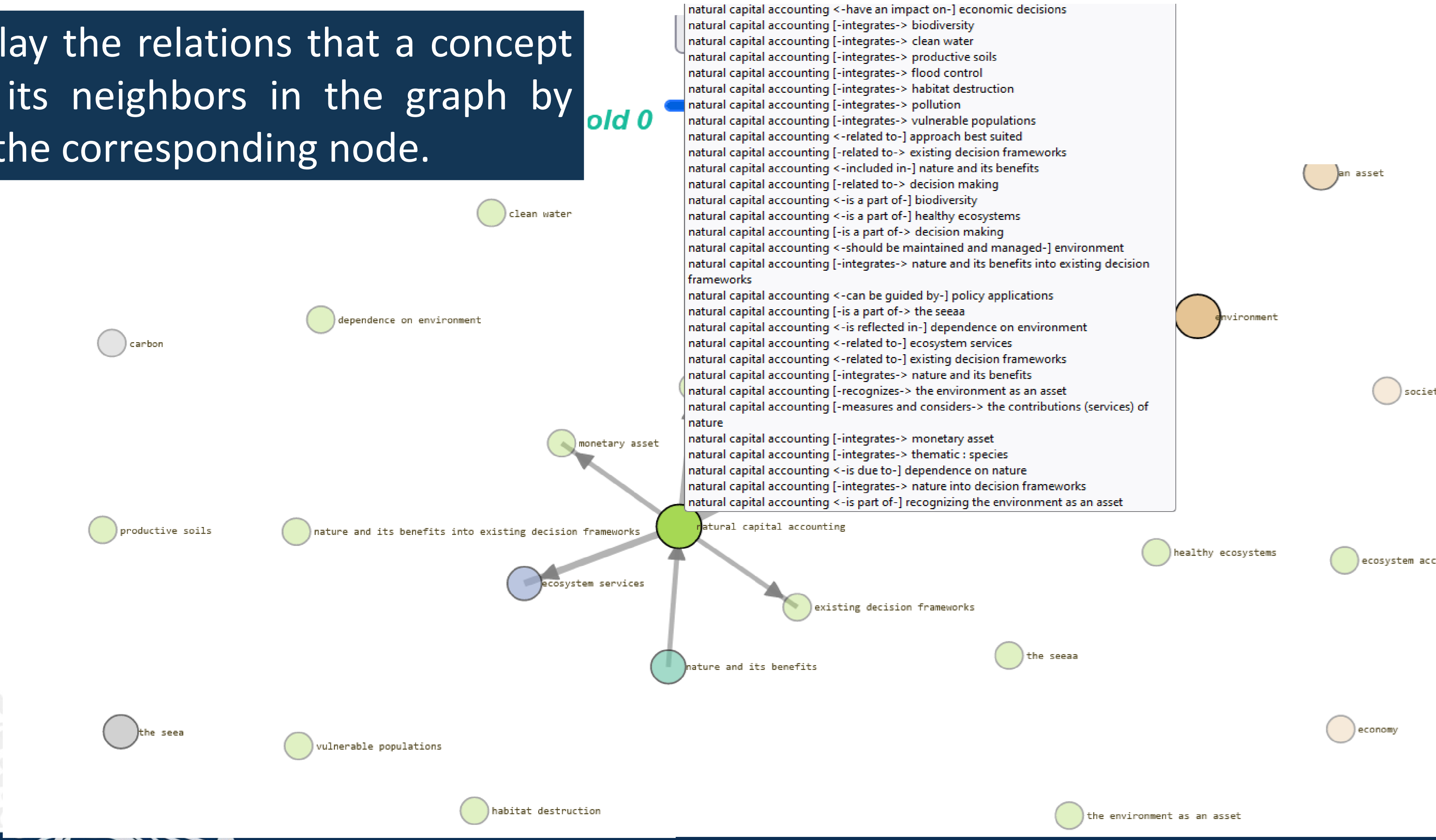
You can choose to view all the links in the (sub) knowledge graph – threshold 0 – or only the most important links (those mentioned the most often in the course document – threshold 1 to 5)

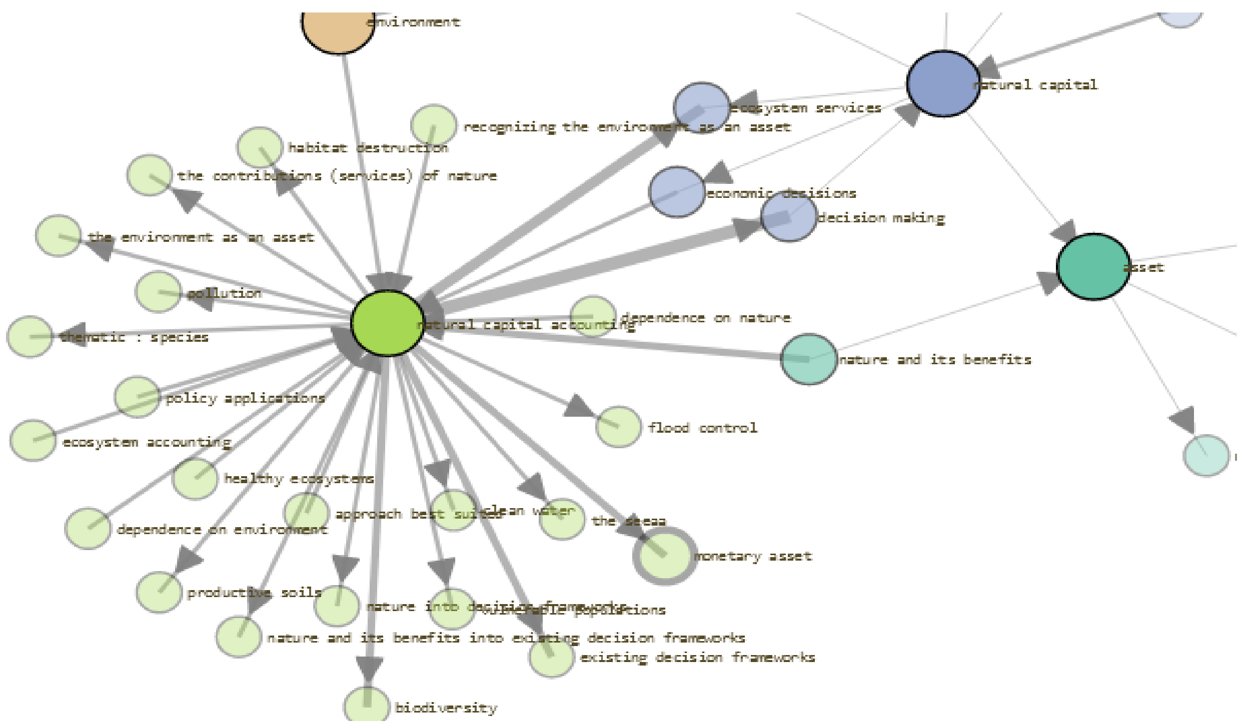


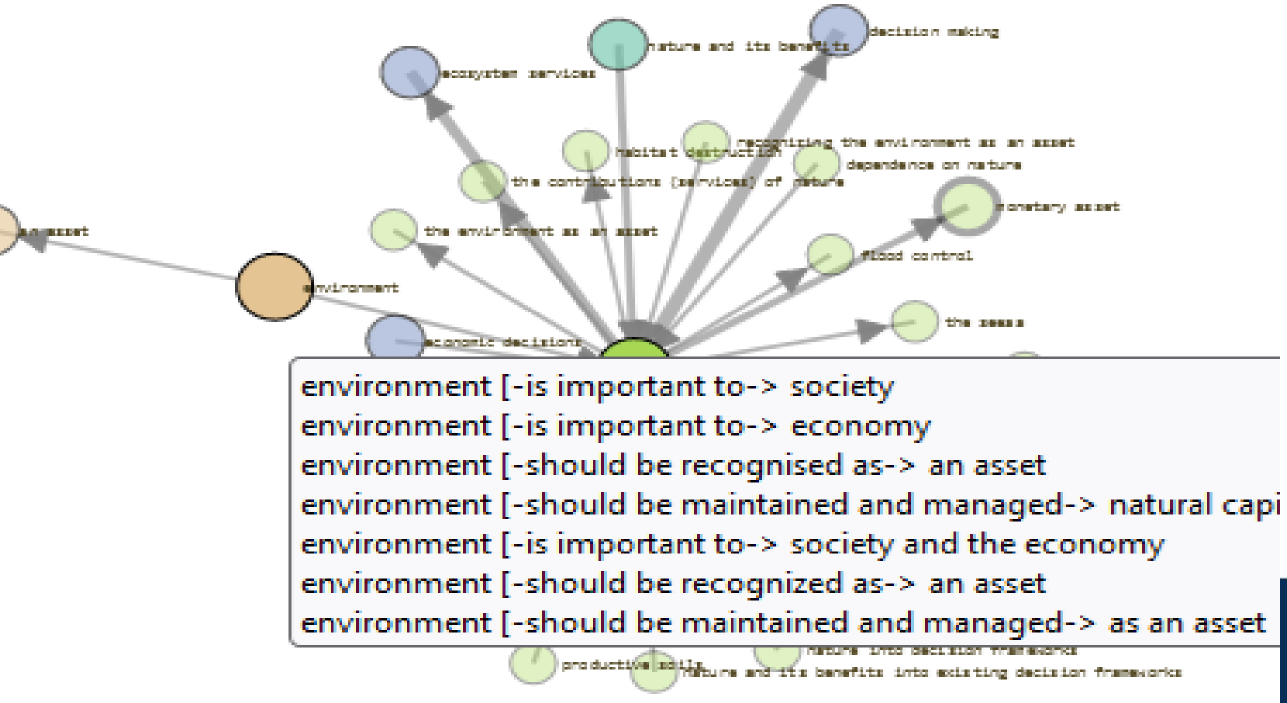


Helping the student navigate in a graph of closely related concepts

You can display the relations that a concept shares with its neighbors in the graph by hovering on the corresponding node.





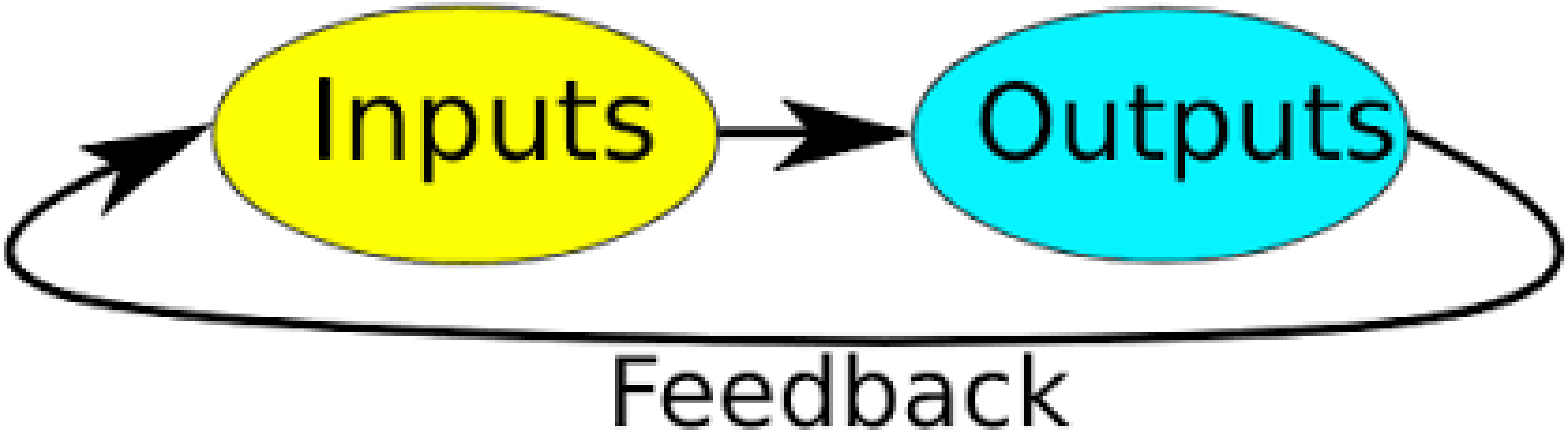


Engaging Through AI-driven Interaction

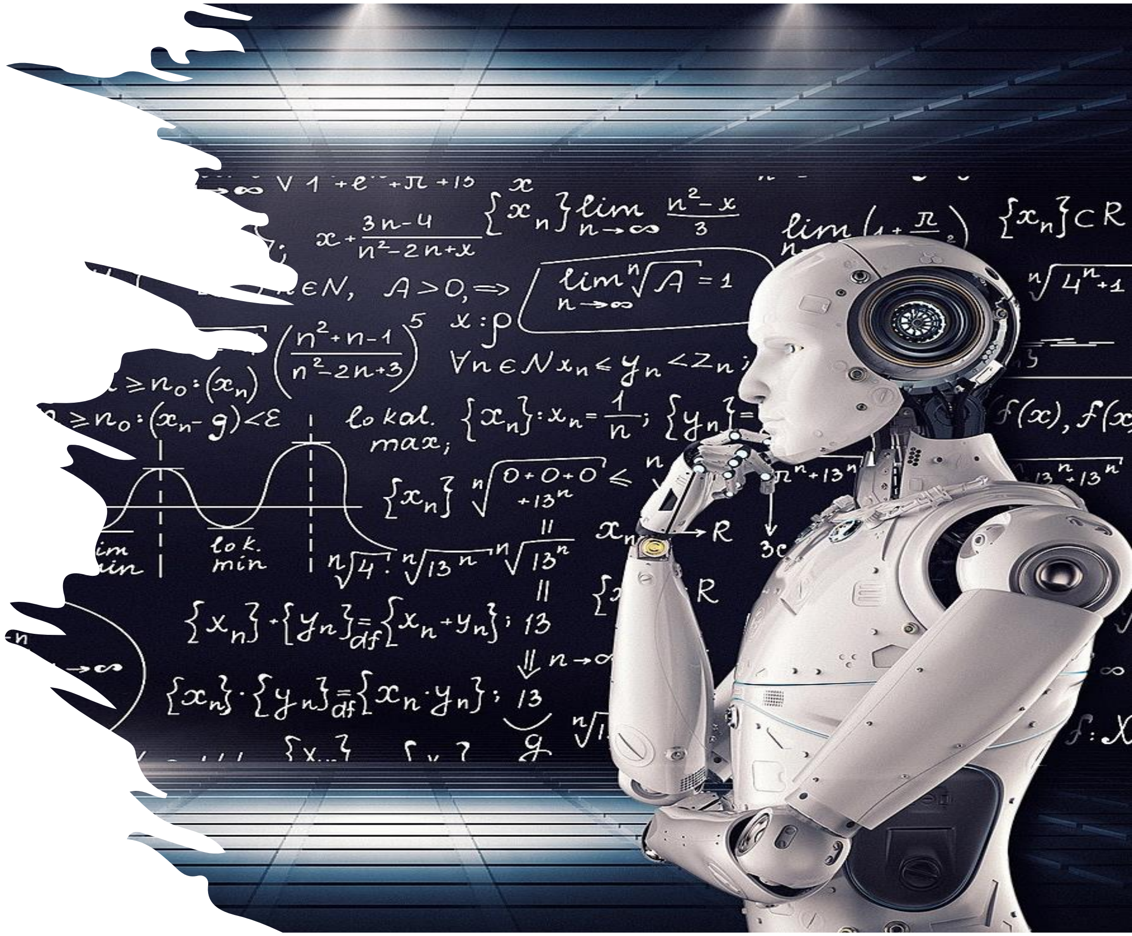
- **Conversational Learning**: Dialogue with AI chatbots for immediate query resolution and concept exploration.
- **Simulations**: AI-powered simulations offering real-life scenarios for hands-on experience.
- **Case Studies**: Interactive case studies integrated with AI insights to apply EEA principles in real-world contexts.
- **Personalized Feedback**: Instant, personalized feedback on exercises and quizzes, enhancing learning outcomes.

Leveraging Feedback for Excellence

- **Participant Ratings**: Collecting participant evaluations to identify strengths and areas for enhancement.
- **Continuous Content Update**: Utilizing feedback to refine course materials and AI responses for relevance and accuracy.
- **Adaptive Learning Models**: AI algorithms adjust based on feedback to improve engagement and understanding.



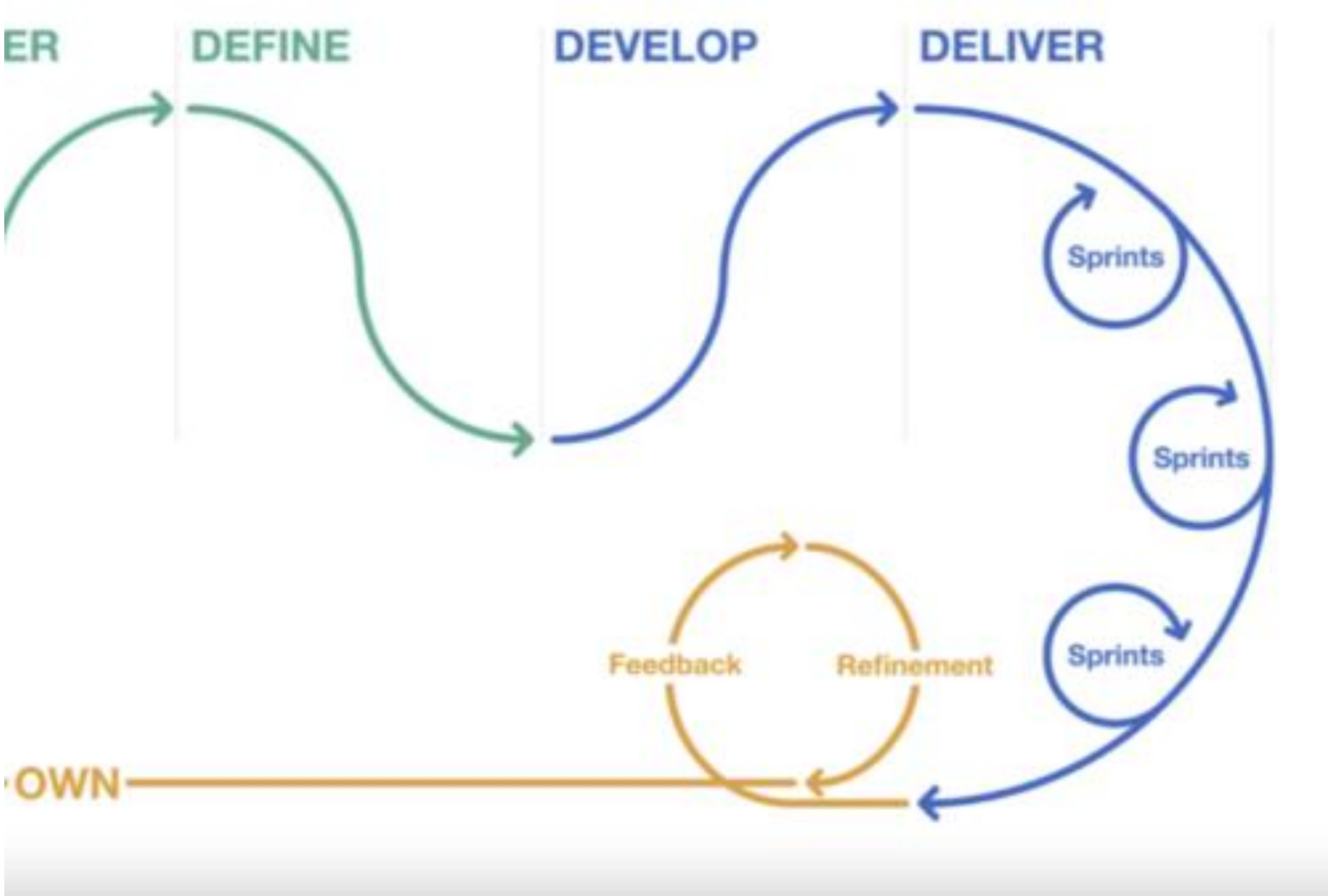
Enhancing Learning with AI



- **Personalized Learning**: AI enables customized learning experiences tailored to individual needs and pace.
- **Immediate Feedback**: Real-time responses and assessments from AI enhance understanding and retention.
- **Engagement and Interaction**: Conversational AI fosters a more interactive and engaging learning environment.
- **Comprehensive Access**: AI consolidates and simplifies access to diverse and complex information sources.
- **Data-Driven Insights**: AI analytics provide insights into learning patterns, helping to optimize the educational journey.

Ensuring Course Effectiveness

- **Feedback Collection**: Regular collection of feedback from participants to gauge satisfaction and areas for improvement.
- **Performance Metrics**: Analysis of completion rates, quiz scores, and engagement levels to assess learning outcomes.
- **AI Analytics**: Utilization of AI-generated data to understand participant interactions and adjust content dynamically.
- **Continuous Improvement Process**: Implementing changes based on evaluations to refine and enhance the course over time.
- **Collaborative knowledge building**: the users can contribute their own answers and use cases to enrich the course content.



Shaping the Future of SEEA Learning

- **Transformative Learning**: AI-assisted learning has reshaped access to and engagement with SEEA.
- **Achievements**: The course had success in making SEEA accessible, interactive, and effective.
- **Future Enhancements**: There are many potential updates and expansions based on evolving AI technologies and participant feedback.
- **Broader Impact**: There are wider implications for educational practice, capacity building in statistics, and sustainable development.

IOS-ISI 2024 MEXICO CONFERENCE

IAOS-ISI 2024
MEXICO CONFERENCE

Thank you

INEGI

isi International Statistical Institute

IAOS IMPROVING OFFICIAL STATISTICS