Weather shocks and migrations Intension in Western in Africa

John Aoga, Juhee Bae, and Stefanija Veljanoska

UCLouvain
icteam
Artificial Intelligence & Algorithms

Our Motivation

Why we are doing this research ?

OBJECTIVE OF THIS STUDY

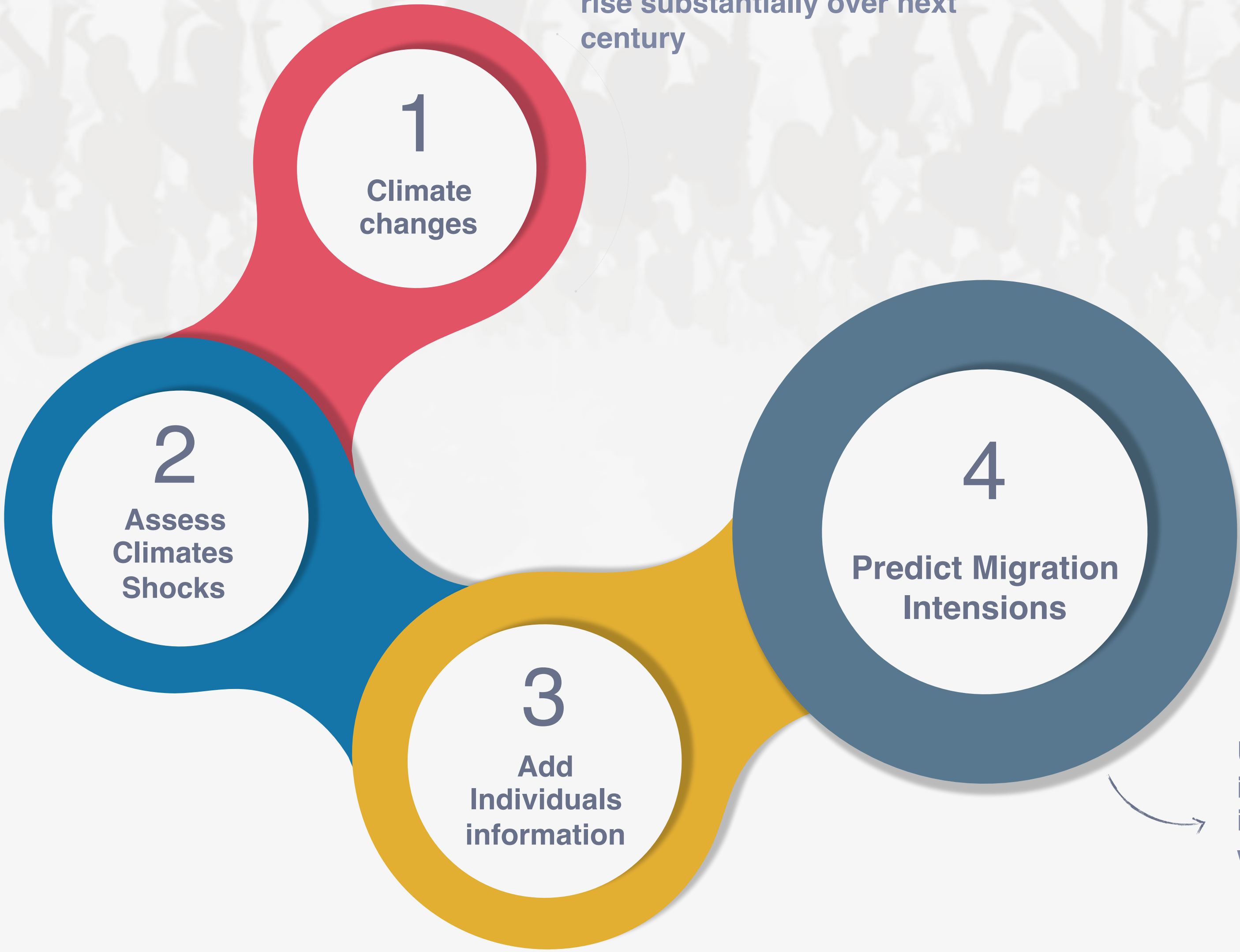
What is our goal?

Understand and interpret migration intensions

HOW?

Weather shocks + individuals information => migration intension

Global temperatures expected to rise substantially over next century



Understand and interpret migration intensions in relation with climate changes

WHY WEATHER?

Combine Economic and climate shocks variables

- Changes in weather conditions induce economic, health and welfare effects within a given spatial unit [1,2].
  - Temperature or Rainfall have strong impacts on agriculture-dependent economies (like west-africa countries).
  - Other parameters can also influence economic outcomes => it is **difficult to identify the causative effects of climate shocks**. [1].

Interesting paper

[1] Dell, M., Jones, B. F., & Olken, B. A. (2014). **What do we learn from the weather? The new climate-economy literature.** *Journal of Economic Literature*, *52*(3), 740-98.
[2] Rigaud, K., Jones, B., Bergmann, J., Clement, V., Ober, K., Schewe, J., Adamo, S., McCusker, B., Heuser, S., and Midgley, A. (2018). Groundswell: Preparing for internal climate migration. Washington, DC: The World Bank.

Research questions

- How can we explain migratory intentions based on climate shocks?

- Which time horizon is mandatory to capture the shocks which impact the decision of people to move?

- Which shocks (variables) most affect people's decisions to move (internally and internationally)?

- Which time horizon is mandatory to capture the shocks which impact the decision of people to move?

WEATHER SHOCKS

Temperature, rainfall, SPEI

Covered in a previous presentation

Multilevel approach

Bertoli, Docquier, Rapoport, and Ruyssen

What is parameter learning ?

y — Migration Var

C — Climate Vars

X — Control Vars

t — Time

$$y = f(C, X, t)$$

Global Model

WE ARE LOOKING FOR f

[1] Bertoli, S., Docquier, F., Rapoport, H., & Ruyssen, I. (2019). Weather shocks and migration intentions in Western Africa: Insights from a multilevel analysis Workshop on Climate change, Inequality and Human Migration, AFD, Oct 2019, Paris, France.

DATA DESCRIPTION

From the Gallup World Polls data (1,7 millions of obs. x 2,600 of vars.)

- Targeted countries (over 9 years, ~60,000 obs. x 900 vars)
  - Burkina Faso, Ivory Coast, Mali, Mauritania, Niger, and Senegal.
    - The most « at risk » regions of the world in term of environmental balance and associated mobility patterns

- Migration intensions => **migration var**
  - Q1 (internal migration). In the next 12 months, are you likely or unlikely to move away from the city or area where you live? **(BMIG_in)**
  - Q2 (international migration). Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country? **(move)**

DATA DESCRIPTION

From the Gallup World Polls data (1,7 millions of obs. x 2,600 of vars.)

- Economic variables => **Control variables.**
  - **Lnhhincpc** => Household size
  - **hhsize** => Household size
  - **children** => Number of children
  - **urban** => Urban/Rural area
  - **mabr** => Connexion abroad (network variable)
  - **hskill** => Education (Highly/ educated
  - **male** => Gender
  - **age1524, age2534 and age35plus** => age variables (intervals [15, 24], [25,34], [35, 45])

In the paper, they showed why these control variables are important

- Temperature and Rainfall from CRU-TS 4.01 gridded datasets => **Climate variables**
  - Compute long-term mean and (Relative and absolute) Standard deviations.
    - Over 36 months.

- Standardized precipitation Evapotranspiration Index (SPEI) => **Climate variables**
  - Drought index used to determining the onset, duration and magnitude of drought conditions
  - It depends on several climate variables such as rainfall, temperature, and evapotranspiration.

https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.01/

ENRICHED DATASET

Gallup dataset + Weather Shocks dataset => join by region

- 39319 obs. x 8441 vars. (only 12 control variables)

- Names of climate variables

T, P, SPEI
Temperature, Rainfall or SPEI over 12 months (alternatives are 1, 3, 6, 18 or 24 months)

r,a
Absolute (a) or relative (r) deviations from the mean

A,B,D
A denotes that all months (regardless of the growing season) have been considered (B indicates only months falling within the planting season were considered; D indicates only months falling within the planting+harvesting season were considered)

D1-D9
They differentiate between the size and direction of anomalies

C1-C36
The period cover by this variable (month-1, month-36)

https://gadm.org/metadata.html
https://www.dropbox.com/sh/b16zbcl6t3uh6v6/AAAjrKc_eJaYwjZzV5sYmJJma?dl=0
https://www.dropbox.com/sh/v8q74qt3uujrb5y/AACFl9wsmhHiEaQ7ECB27H2ta?dl=0

REGRESSIONS AND CONCLUSIONS

310 000 logit regressions

- 310 000 logit regressions (= 6 x 7 x 3 x 3 x 36 x 2 x 2 x 7)
  - Six countries
  - Seven weather variables of interest (T, P, 5 SPEIs)
  - Three types of weather shocks
  - Three measures of the intensity of the shocks (1, 2, 3 deviations)
  - 36 months
  - Two anomaly period specifications (all months or only crop-growing season)
  - Two types of regional identifiers (finer or coarser)
  - Seven samples (full sample, urban/rural areas, low/high-educated respondents, with/without connection abroad)

- The predictive power is maximized when :
  - using **negative SPEI** shocks (i.e., droughts),
  - measuring shocks as the share of months with **at least 2 relative standard deviations** below the local SPEI long-term value **over the last 12 months**,
  - focusing on the **crop-growing season** (not all months)
  - focusing on the subsample of **individuals living in rural areas**.

Considering the crop-growing season over the previous 12 months

- Analysis for **people migration**
  - higher probability of intending to move for **Senegal, Niger, and Ivory Coast**.
  - Insignificant for the other countries.
- Analysis for **international migration**
  - higher probability of intending to move from **Niger** only.

Our Approach

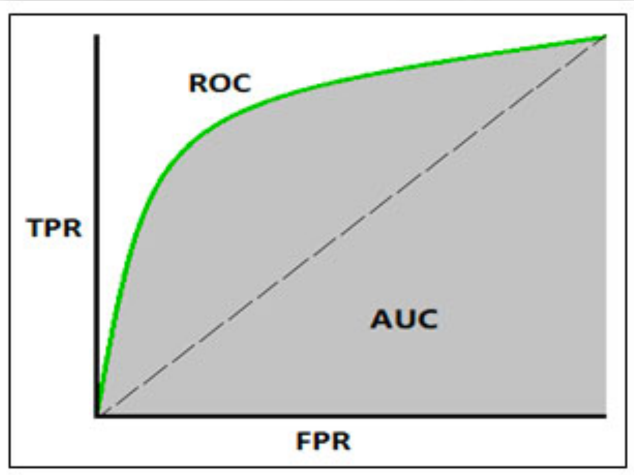
New lines of research (machine learning)

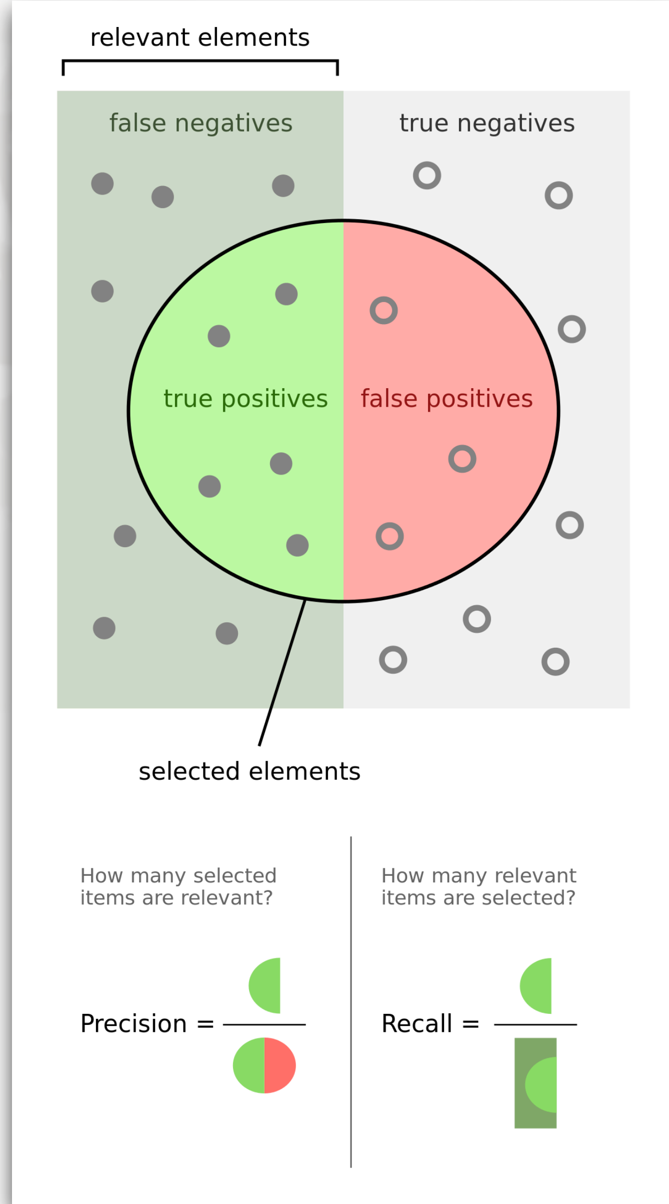
MACHINE LEARNING APPROACHES

Differences with manual regression

- Splitting data into **train and test sets** and train set into train and cross validation sets (avoiding overfitting, robustness against noise).

- Measuring the performance of the learning using metrics such as **Precision, Recall, Accuracy, F-measure and the Area Under the curve (AUC)**.





- Confirming the result of the paper ? Drawing further conclusions.
  - Using the **Enriched dataset** (the same as the paper).
  - Preprocessing the enriched data into a **Raw dataset** (i.e. we have the real SPEI values not a ratio).
  - **Visualizing** the dataset in many ways (find correlations, dependencies, outliers, hints, ...)
  - Running several **machine learning algorithm** (mainly tree based method which are really interpretable).
  - **Interpreting** the results (i.e. finding the important features) and validating the results with experts.

Visualization and Interpretation

Check this link

Using Enriched dataset

Run machine learning algorithms to verify the paper conclusions

ENRICHED DATASET

Gallup dataset + Weather Shocks dataset => join by region

39319 obs. x 8441 vars. (only 12 control variables)

Names of climate variables

SPEI12rDD3C20

TaAD3C15

PrBD8C5

T, P, SPEI
Temperature, Rainfall or SPEI over 12 months (alternatives are 1, 3, 6, 18 or 24 months)

r,a
Absolute (a) or relative (r) deviations from the mean

A,B,D
A denotes that all months (regardless of the growing season) have been considered (B indicates only months falling within the planting season were considered; D indicates only months falling within the planting+harvesting season were considered)

D1-D9
They differentiate between the size and direction of anomalies

C1-C36
The period cover by this variable (month-1, month-36)

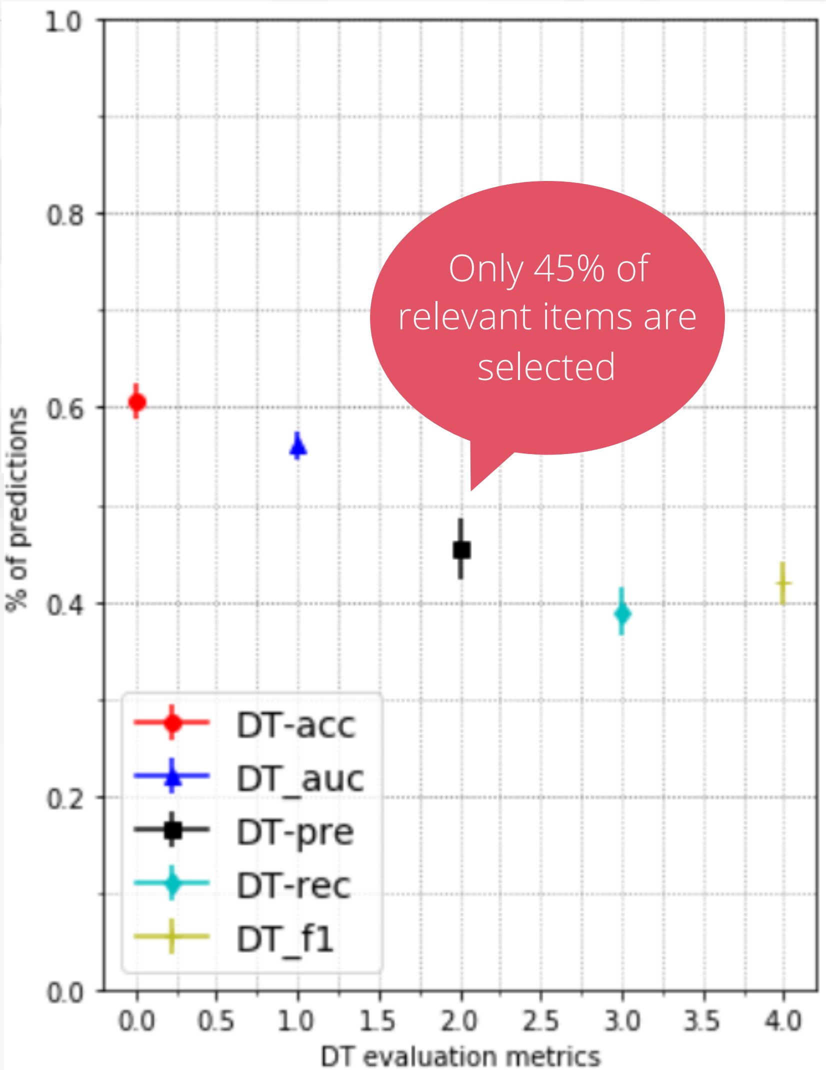
https://www.dropbox.com/sh/b16zbcl6t3uh6v6/AAAjrKc_eJaYwjZzV5sYmJJma?dl=0
https://www.dropbox.com/sh/v8q74qt3uujrb5y/AACFl9wsmhHiEaQ7ECB27H2ta?dl=0

VARIABLES USED FOR SHAP (TRIAL 1)

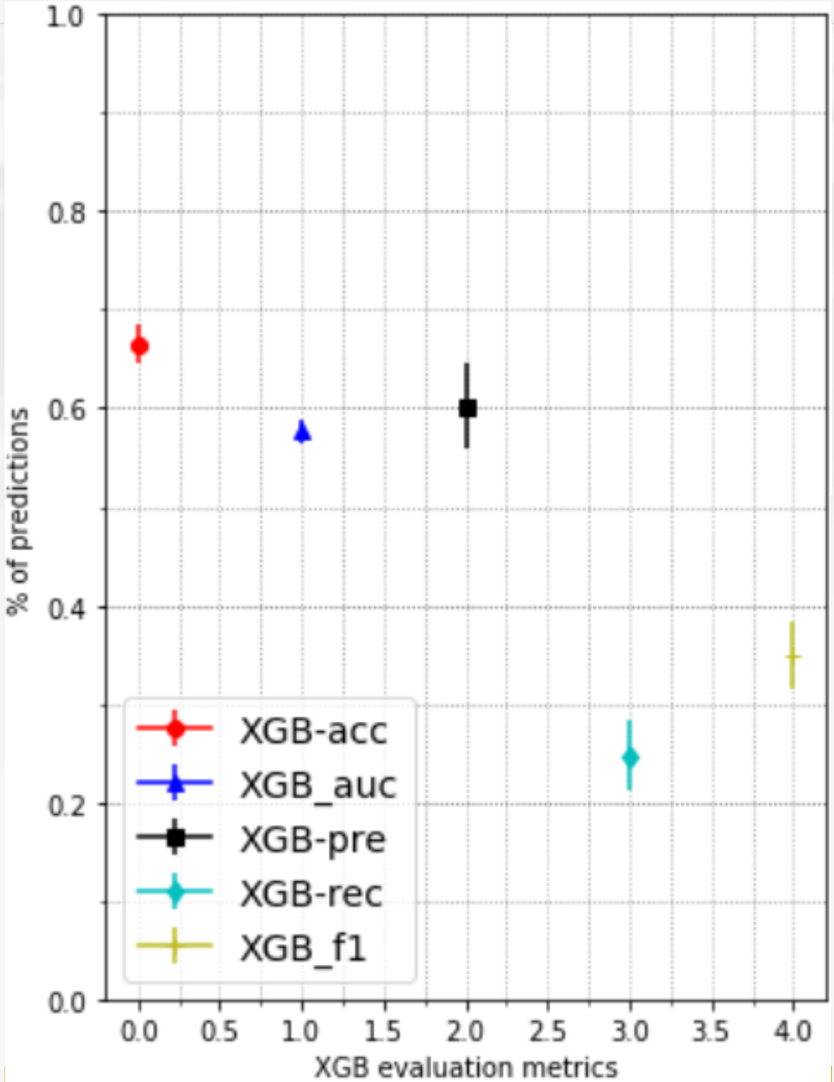
SHAP solution (take two days)

HIGH SIGNIFICANT

LOW SIGNIFICANT



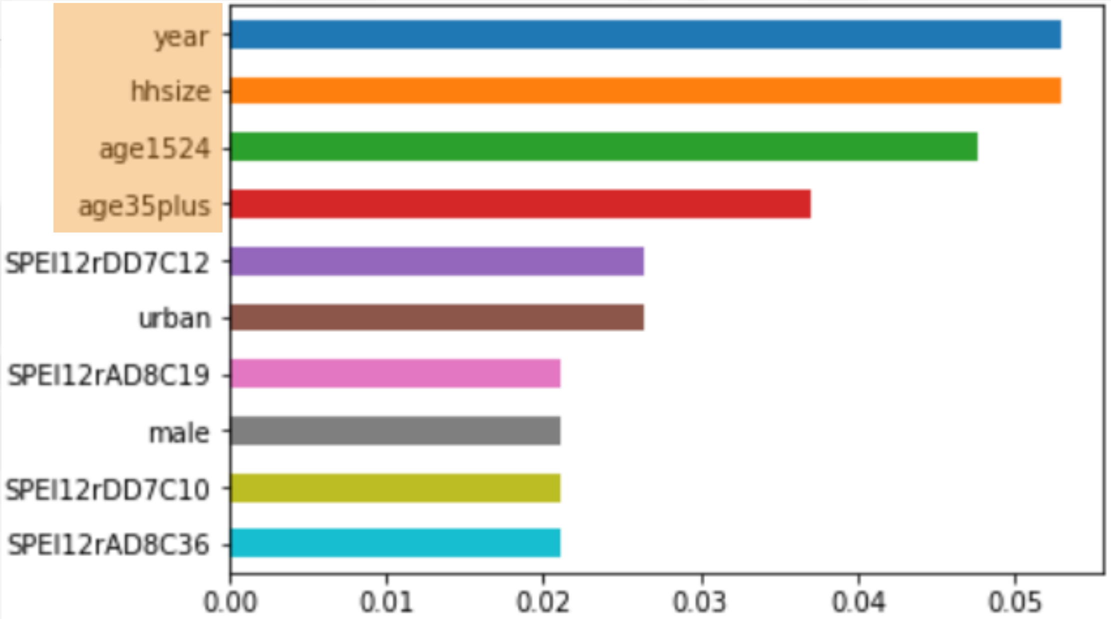
ORIGIN



HHSIZE

PERFORMANCE OF XGB

Feature importance



Feature importance

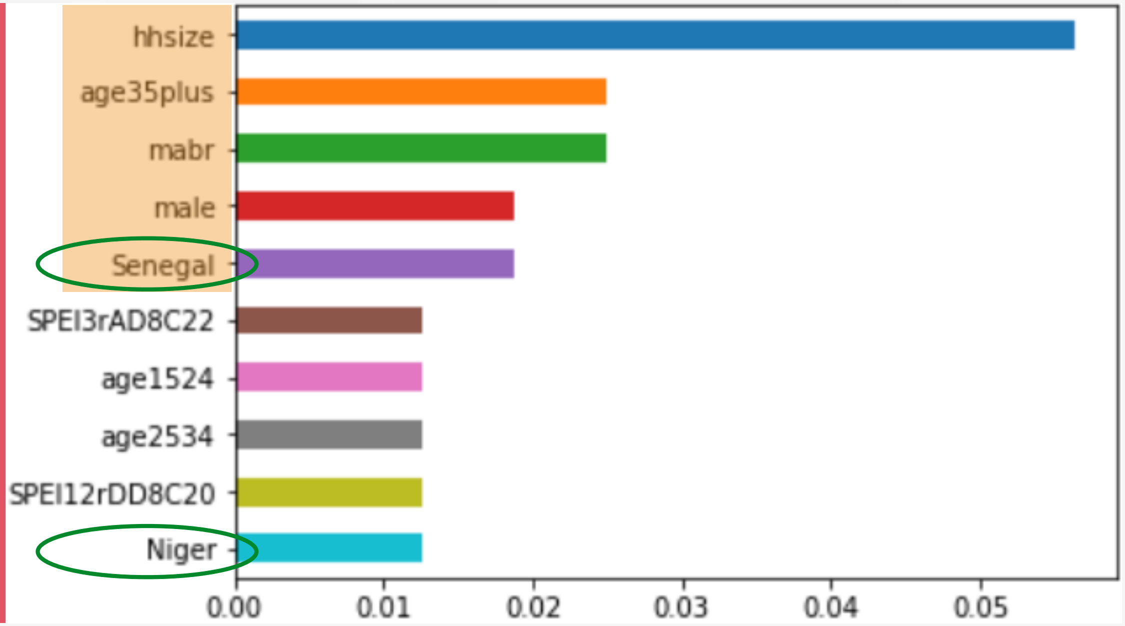
Xgboost: Global feature importance metrics (Senegal)

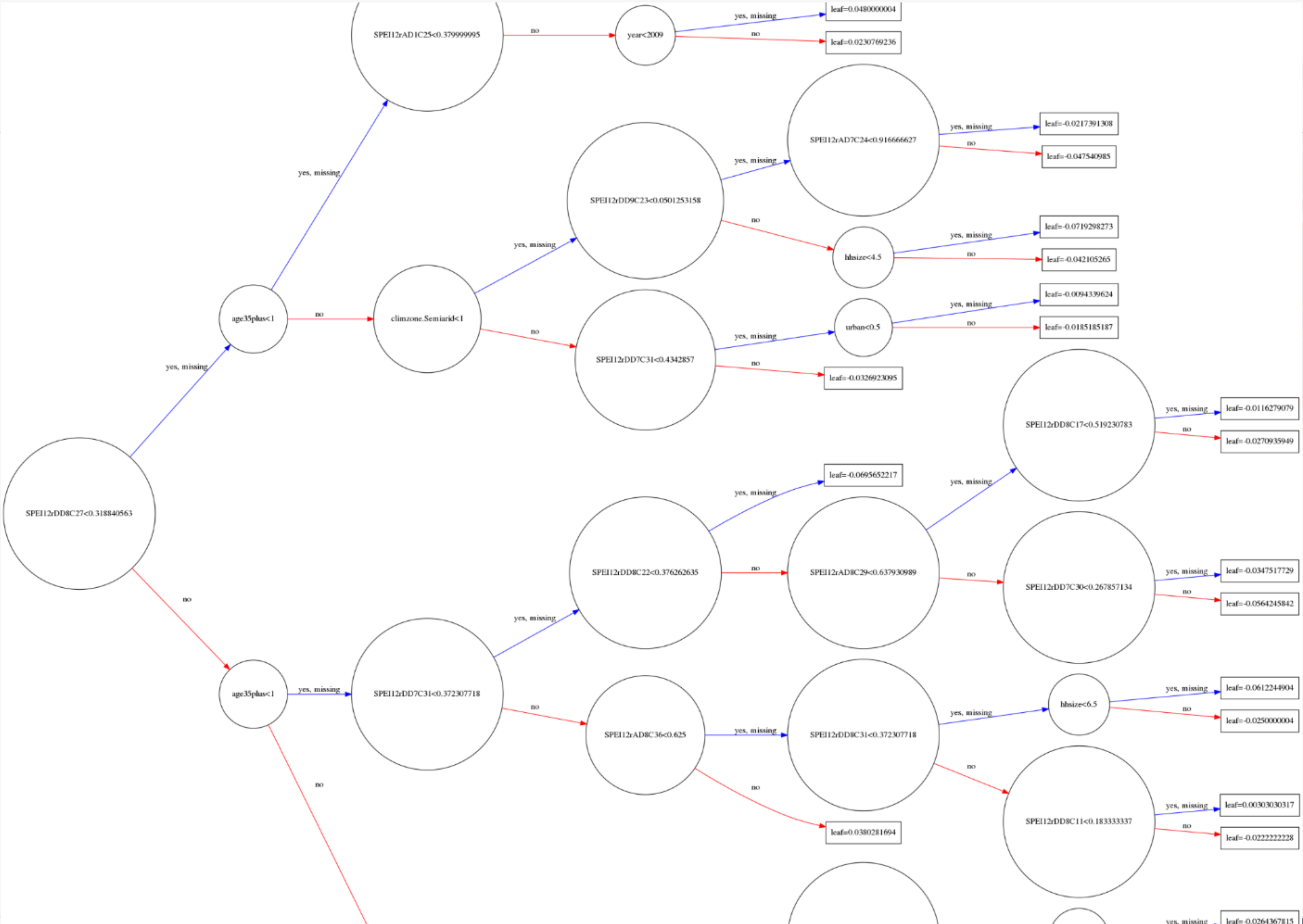
year'	1
Household (hhsize)	2

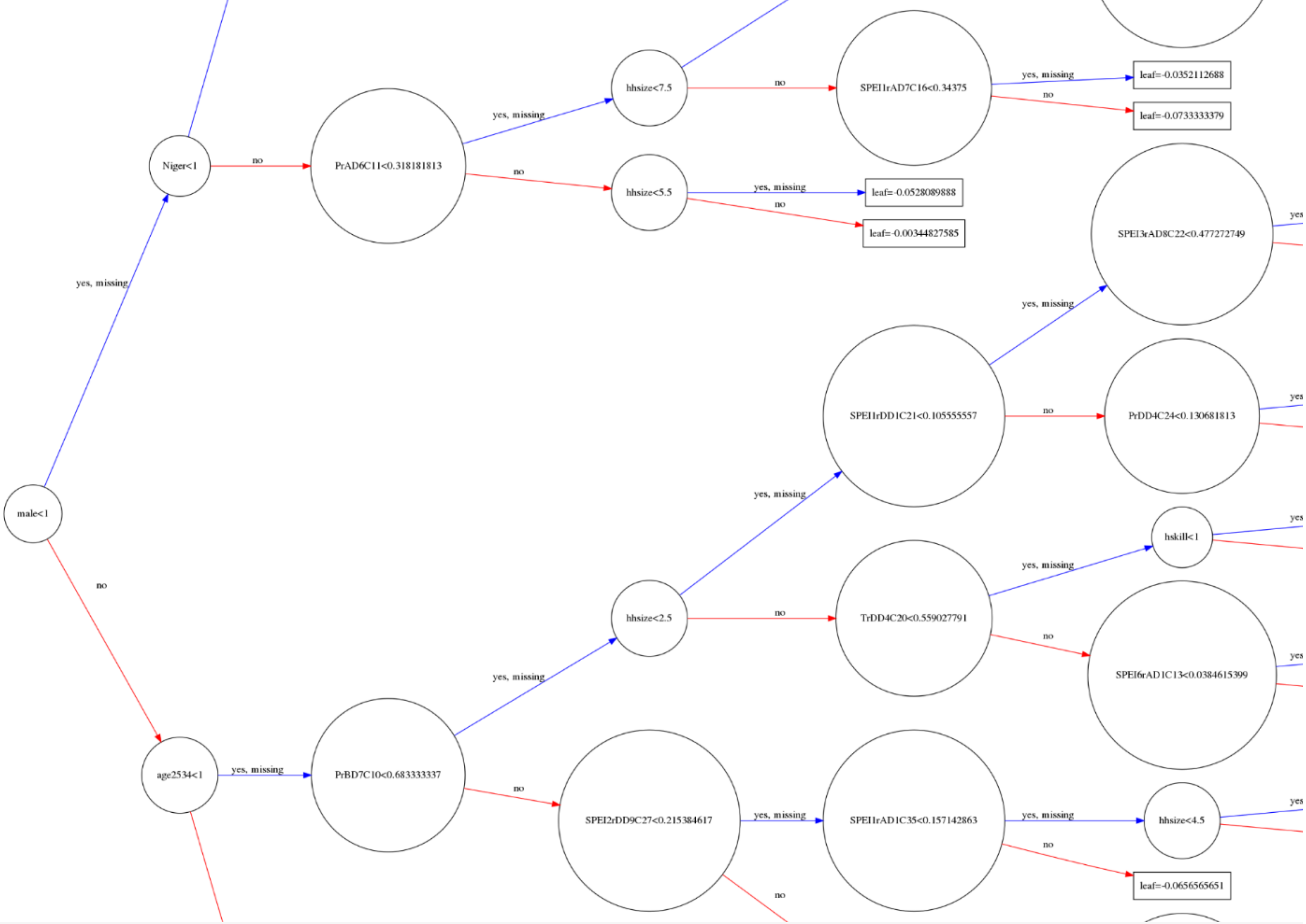
Feature importance

Xgboost: Global feature importance metrics (All countries) - removed age > 35plus no move

Household (hhsize)	1
One-distance Connection Abroad (mabr)	1



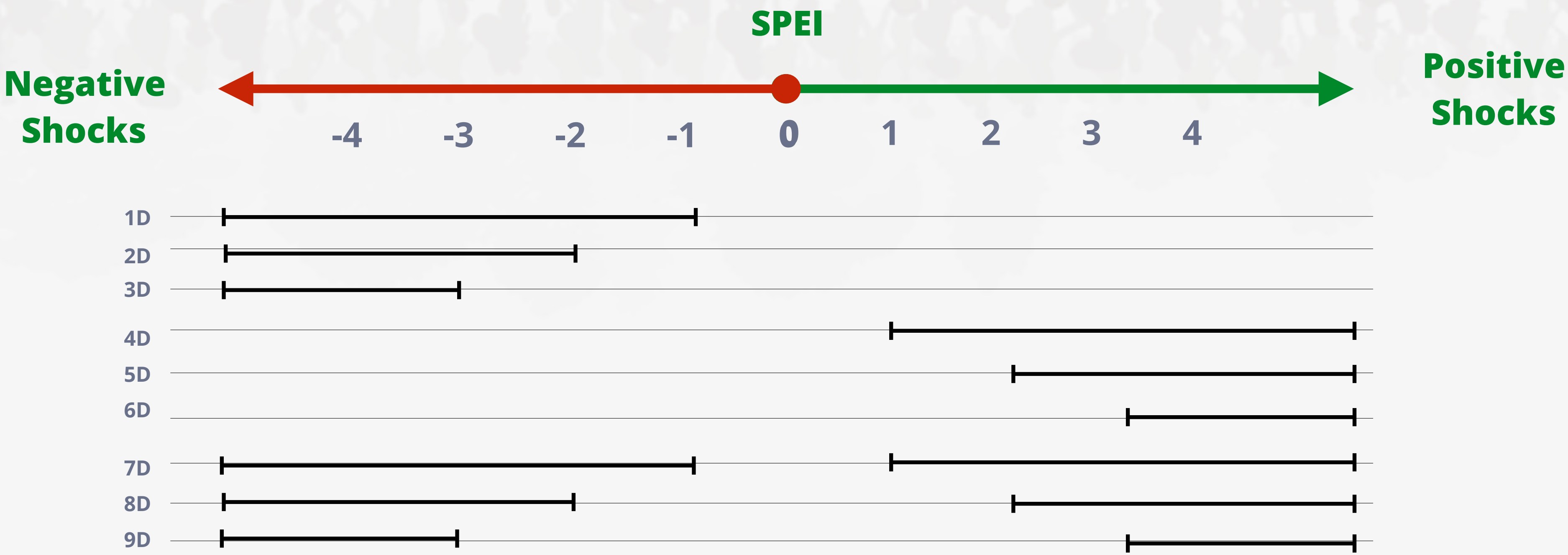

SPEI12rDD8C27<0.318840563
yes, missing
no
age35plus<1
age35plus<1
SPEI12rAD1C25<0.379999995
climzone.Semiarid<1
SPEI12rDD7C31<0.372307718
year<2009
leaf=0.0480000004
leaf=0.0230769236
SPEI12rDD9C23<0.0501253158
SPEI12rAD7C24<0.916666627
leaf=-0.0217391308
leaf=-0.047540985
hhsize<4.5
leaf=-0.0719298273
leaf=-0.042105265
SPEI12rDD7C31<0.4342857
urban<0.5
leaf=-0.0094339624
leaf=-0.0185185187
leaf=-0.0326923095
SPEI12rDD8C22<0.376262635
leaf=-0.0695652217
SPEI12rAD8C29<0.637930989
SPEI12rDD8C17<0.519230783
leaf=-0.0116279079
leaf=-0.0270935949
SPEI12rDD7C30<0.267857134
leaf=-0.0347517729
leaf=-0.0564245842
SPEI12rAD8C36<0.625
SPEI12rDD8C31<0.372307718
hhsize<6.5
leaf=-0.0612244904
leaf=-0.0250000004
SPEI12rDD8C11<0.183333337
leaf=0.00303030317
leaf=-0.0222222228
leaf=0.0380281694
leaf=-0.0264367815


male<1
yes, missing
no
Niger<1
no
PrAD6C11<0.318181813
yes, missing
no
hhsize<7.5
no
SPEI1rAD7C16<0.34375
yes, missing
no
leaf=-0.0352112688
leaf=-0.0733333379
hhsize<5.5
yes, missing
no
leaf=-0.0528089888
leaf=-0.00344827585
SPEI3rAD8C22<0.477272749
yes
yes, missing
SPEI1rDD1C21<0.105555557
no
PrDD4C24<0.130681813
yes
yes, missing
hskill<1
yes
age2534<1
yes, missing
PrBD7C10<0.683333337
yes, missing
hhsize<2.5
no
TrDD4C20<0.559027791
yes, missing
no
SPEI6rAD1C13<0.0384615399
yes
no
SPEI2rDD9C27<0.215384617
yes, missing
SPEI1rAD1C35<0.157142863
yes, missing
hhsize<4.5
yes
no
leaf=-0.0656565651
no

Build new Algorithms from raw data

Running deep learning approaches (RNN)

- 39319 obs. x 8441 vars. (only 12 control variables)
  - 12 Gallup Variables (Same as in the enriched data)
  - Climate Variables (Binarize SPEI + cropping and growing seasons)

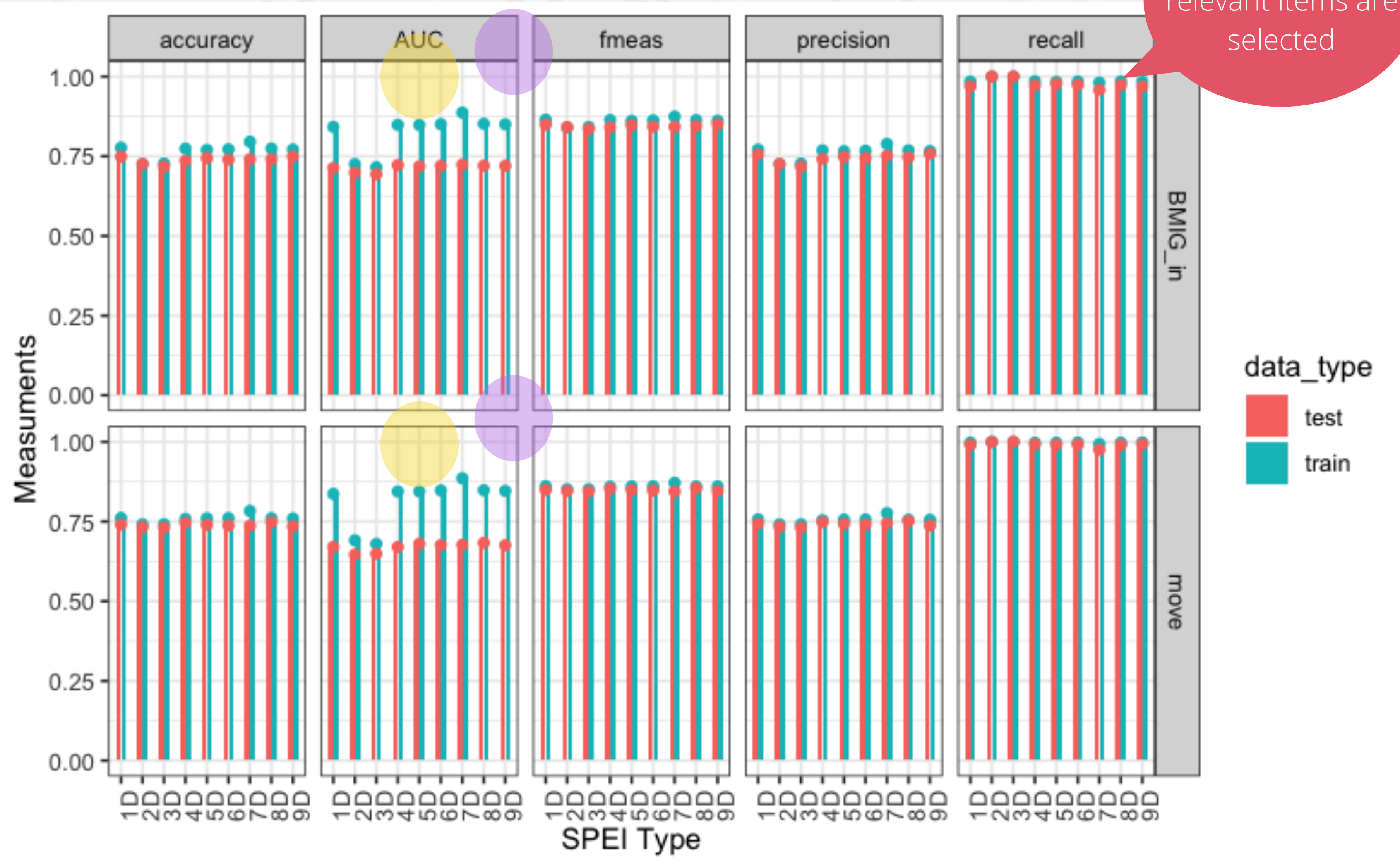


  - 48 Climate variables monthly lags

Performances

PERFORMANCE USING RANDOM FOREST (RF)

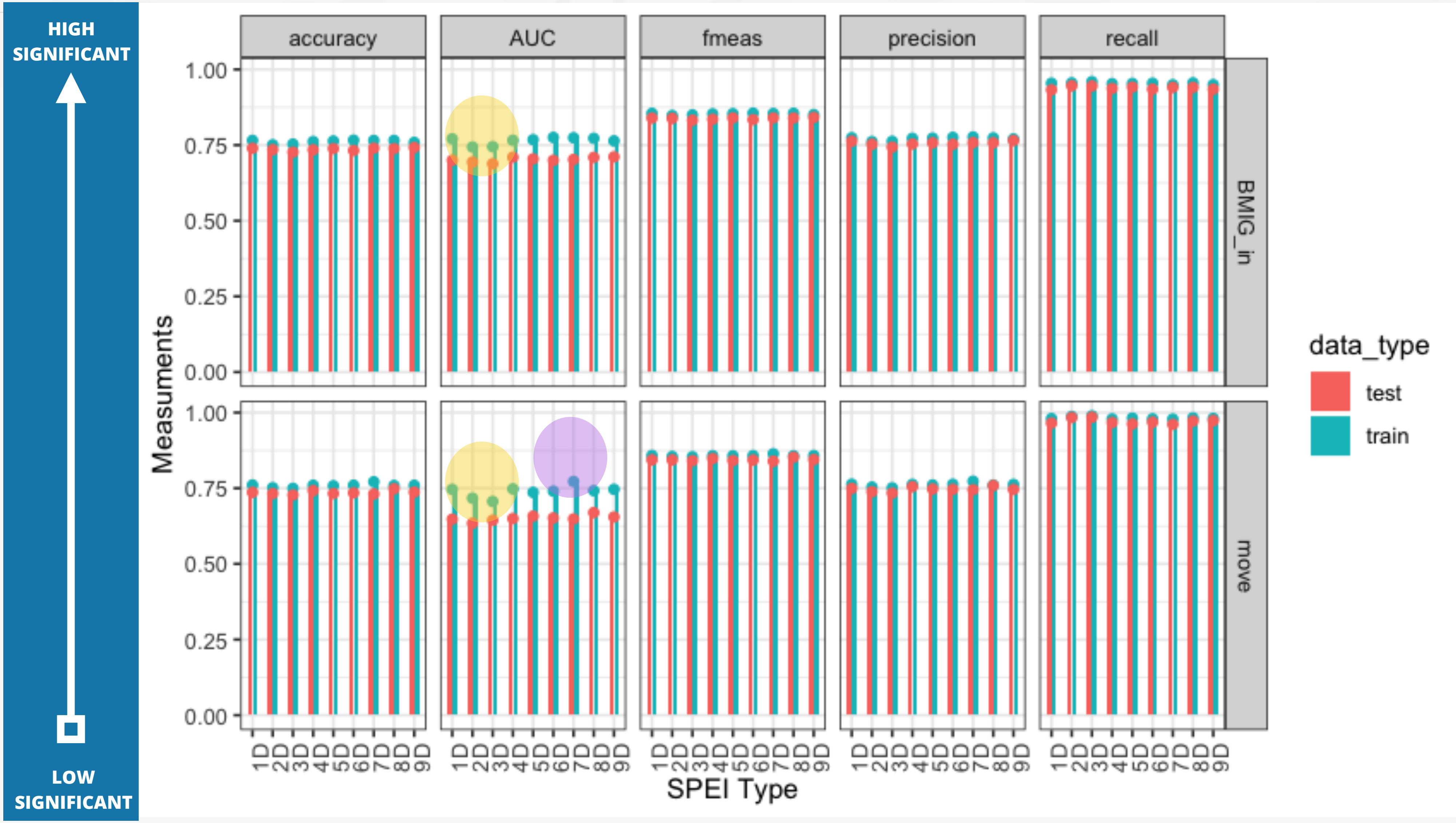
Accuracie and AUC are higher than the other when we have shocks



Almost all of the relevant items are selected

PERFORMANCE USING GRANDIENT BOOSTED TREES (XGB)

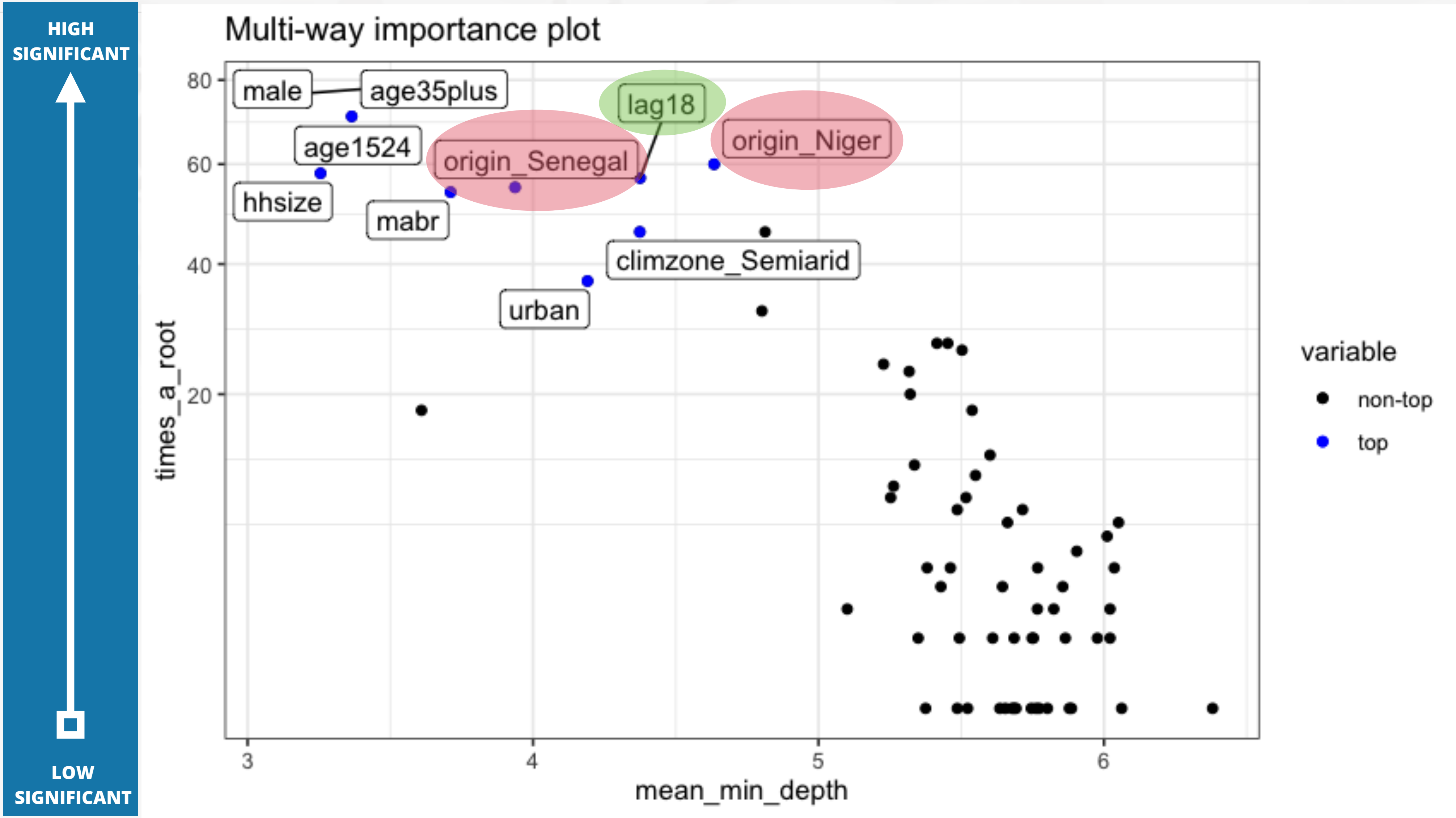
Accuracie and AUC are higher than the other when we have shocks



International Move

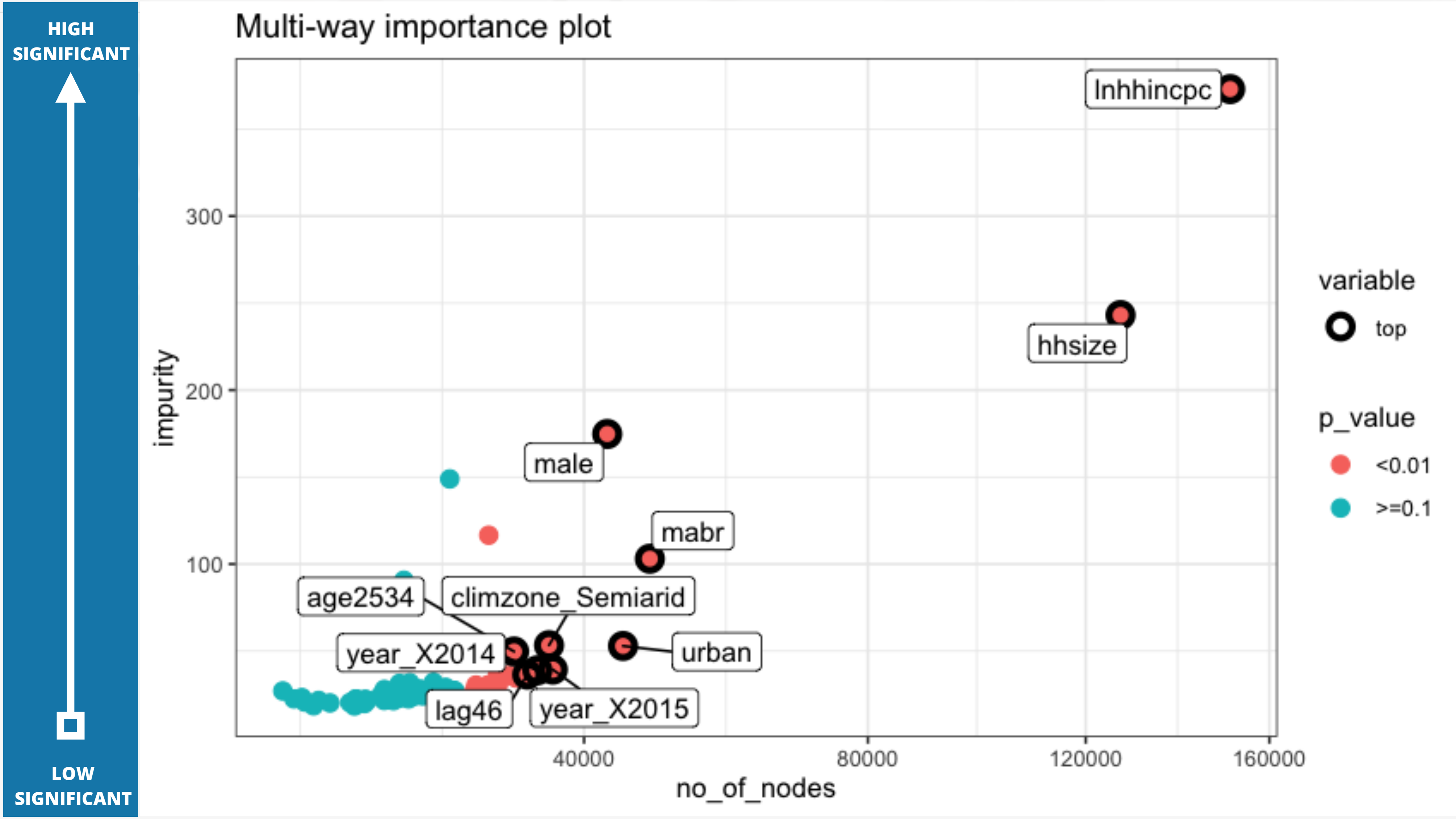
INVESTIGATION OF THE RF MODEL (INTERNATIONAL MOVE)

Feature importance



INVESTIGATION OF THE RF MODEL

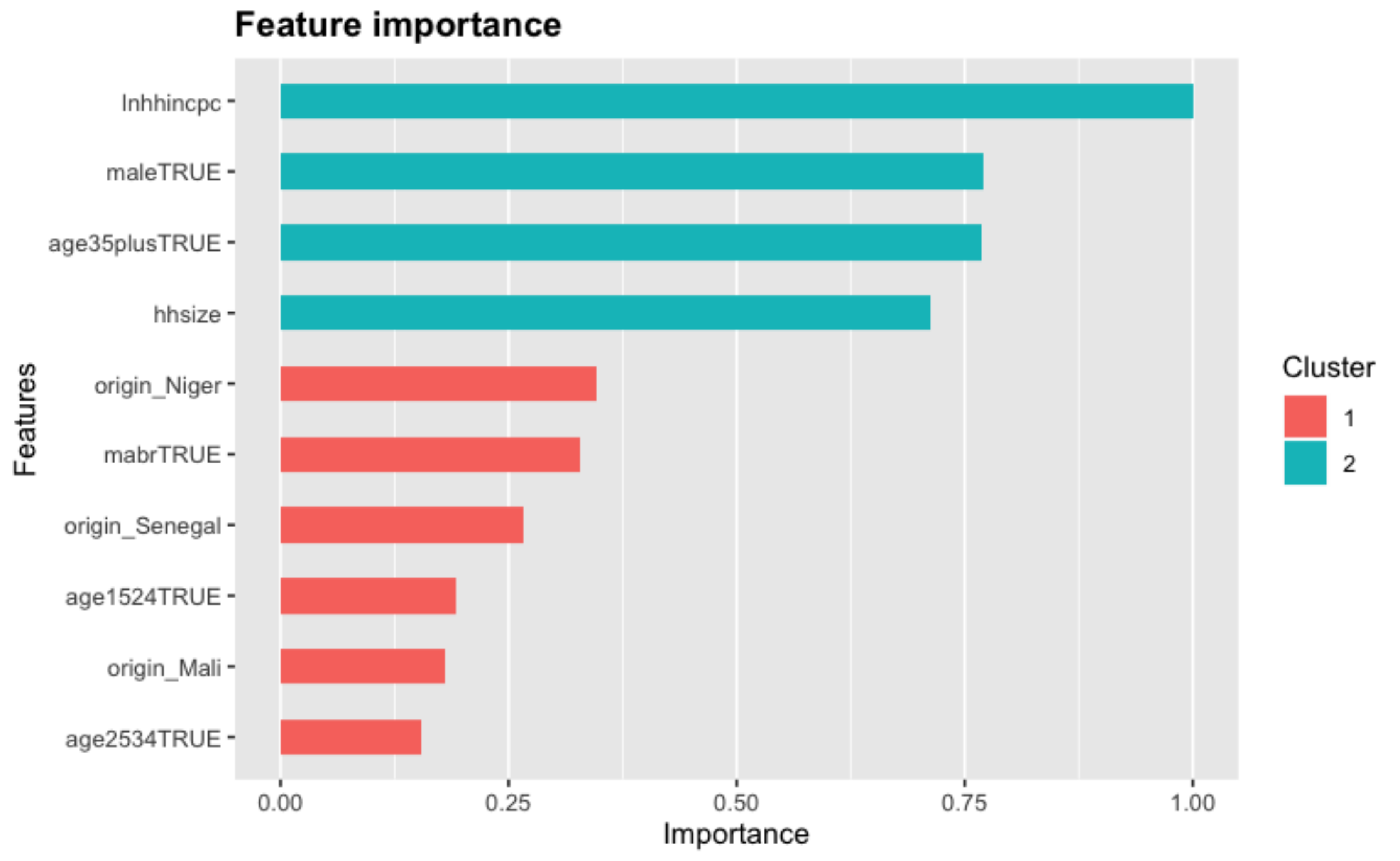
Feature importance (level of noise)



INVESTIGATION OF THE XGB MODEL (INTERNATIONAL MOVE)

Features importance (spei < -1)

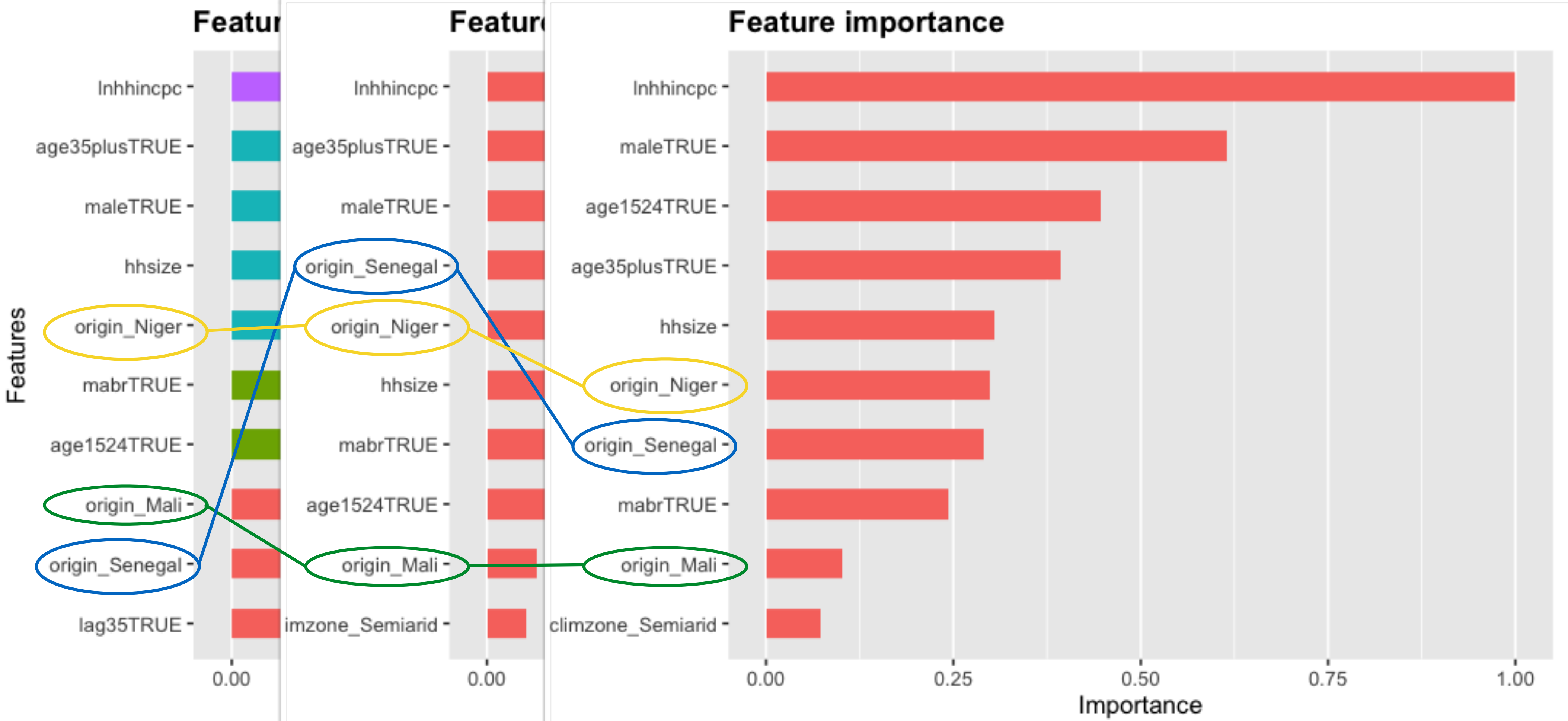
HIGH SIGNIFICANT



LOW SIGNIFICANT

INVESTIGATION OF THE XGB MODEL (INTERNATIONAL MOVE)

Features importance (severe condition)



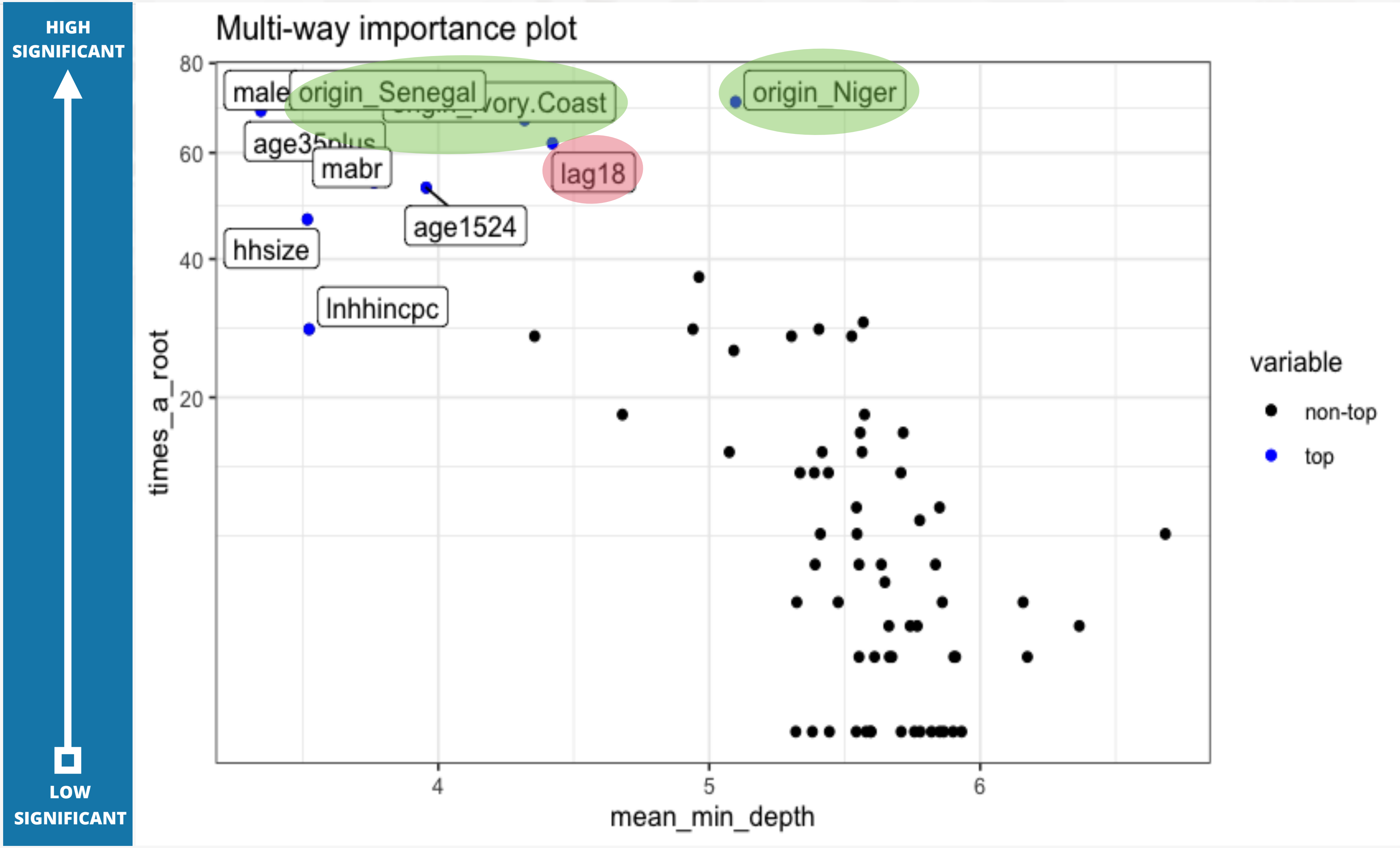
Condition Severity

LOW SIGNIFICANT → **HIGH SIGNIFICANT**

Move

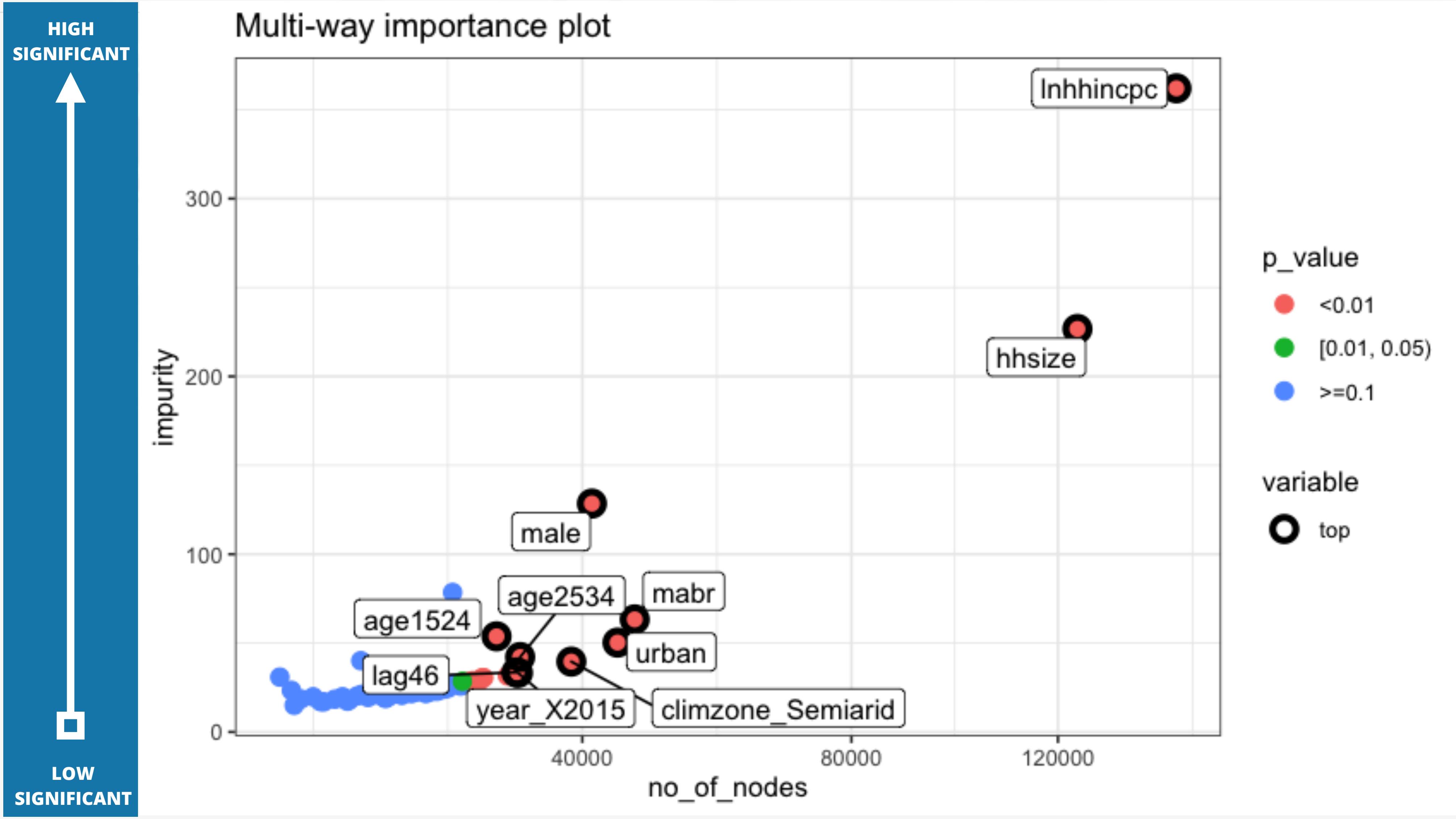
INVESTIGATION OF THE RF MODEL (MOVE)

Feature importance



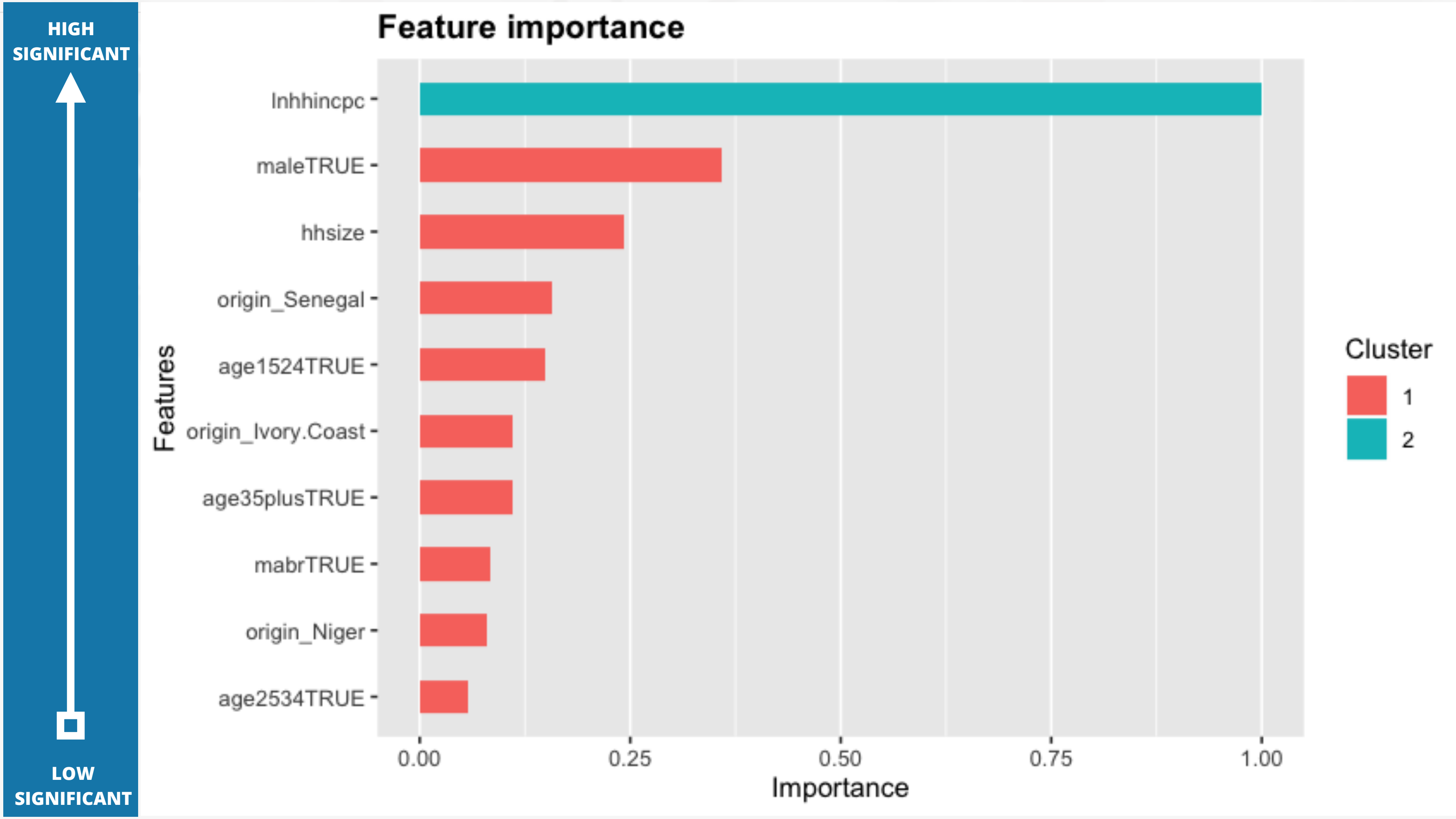
INVESTIGATION OF THE RF MODEL

Feature importance (level of noise)



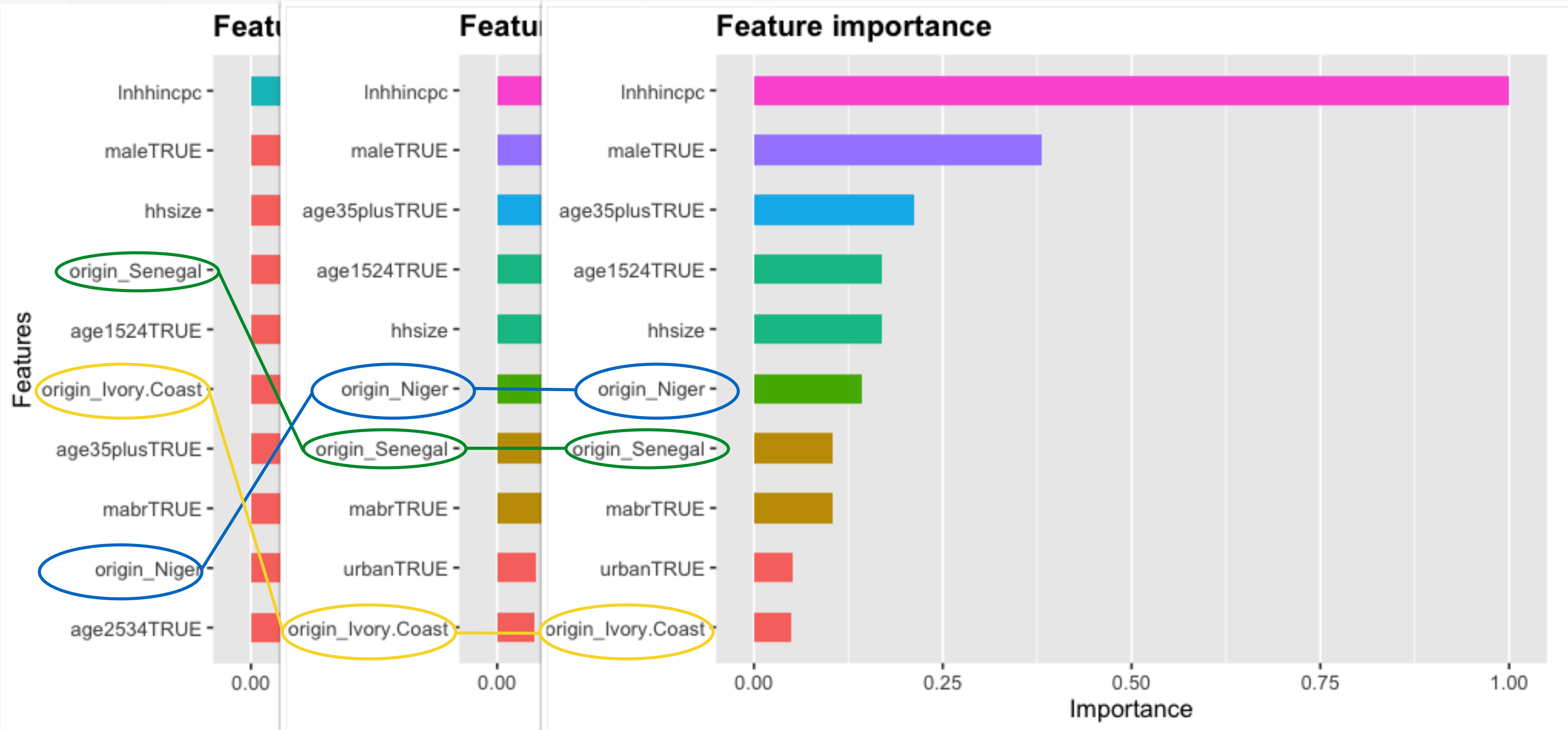
INVESTIGATION OF THE XGB MODEL

Features importance (spei < -1)



INVESTIGATION OF THE XGB MODEL (MOVE)

Features importance (severe condition)



Condition Severity

LOW SIGNIFICANT → **HIGH SIGNIFICANT**

Considering the crop-growing season over the previous 12 months (SPEI)

Findings

- **Income, Gender, Household size and age** play an important role in the people intension to move.

- **18 months** before the interview month is the only interesting weather shocks feature that matters (**=> Difficult to establish a clear link between migration and climate here**)

- **At least 1 degree/stdv above or below LR mean SPEI** affects the people migration intensions.

Considering the crop-growing season over the previous 12 months (SPEI)

Same conclusion as in the paper

- Analysis for **people migration**
  - Higher probability of intending to move for **Senegal, Niger, and Ivory Coast**.
  - Insignificant for the other countries.

- Analysis for **international migration**
  - higher probability of intending to move from **Niger and Senegal in the semi-arid climate zone.**

Next Steps

Time Series + Deep learning approaches (RNN)

Weather shocks and migrations Intension in Western in Africa

John Aoga, Juhee Bae, and Stefanija Veljanoska

UCLouvain icteam Artificial Intelligence & Algorithms

Extra Slides