

Analysis of the impact of socio-economic factors on primary school final examination scores using the SHAP method

Mateusz Bajorek, Zofia Hendrysiak, Aknur Shakhidani, Marius Sorin Pop



Agenda

- Our Team
- SHAP method
- Statistics in Poland
- Gathered data
- Data preparation
- SHAP results for AutoAI model
- SHAP results for Pytorch model
- Prediction error extremes
- Website
- Final thoughts
- References
- Appendix

Our Team



Mateusz



Zofia



Aknur



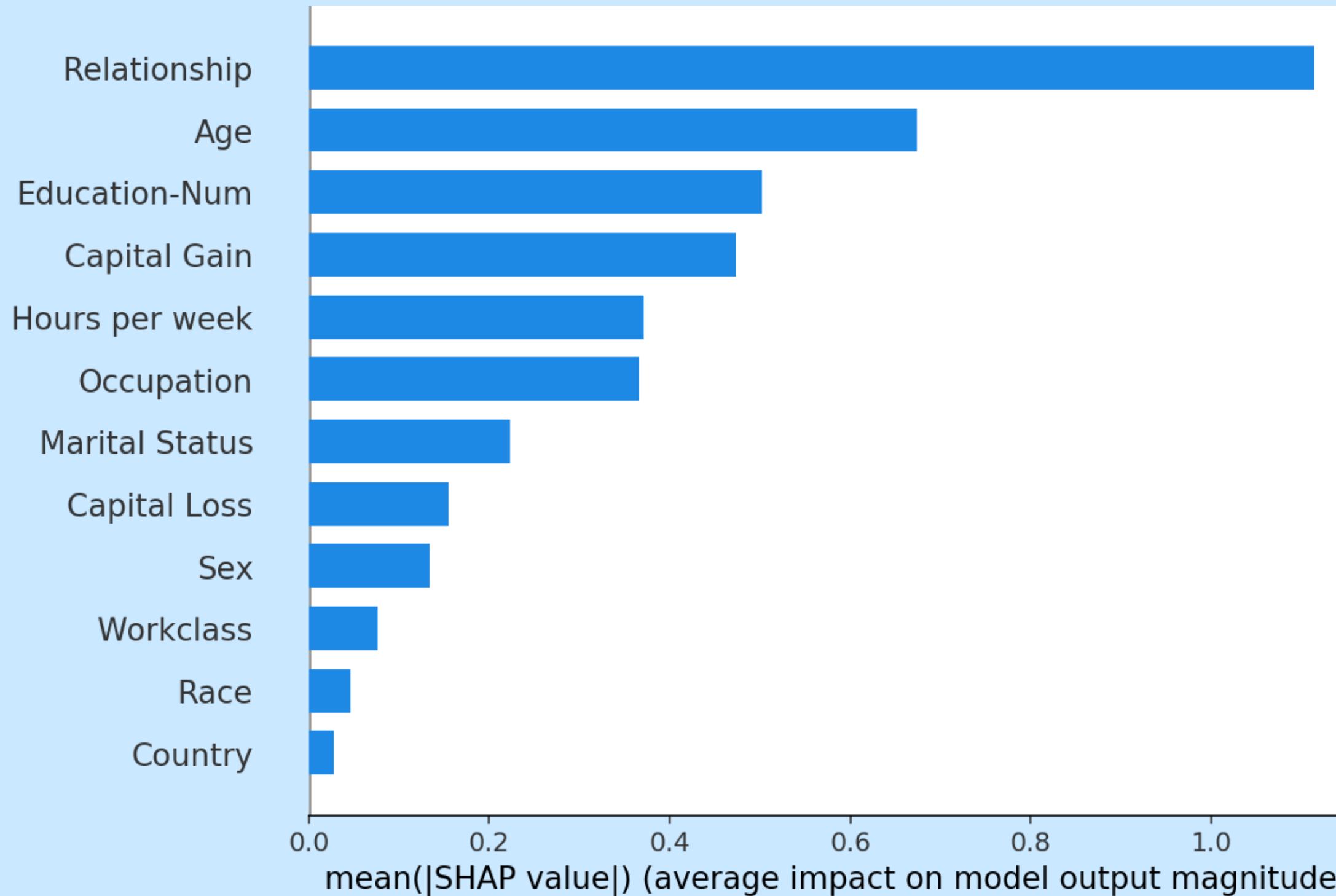
Marius



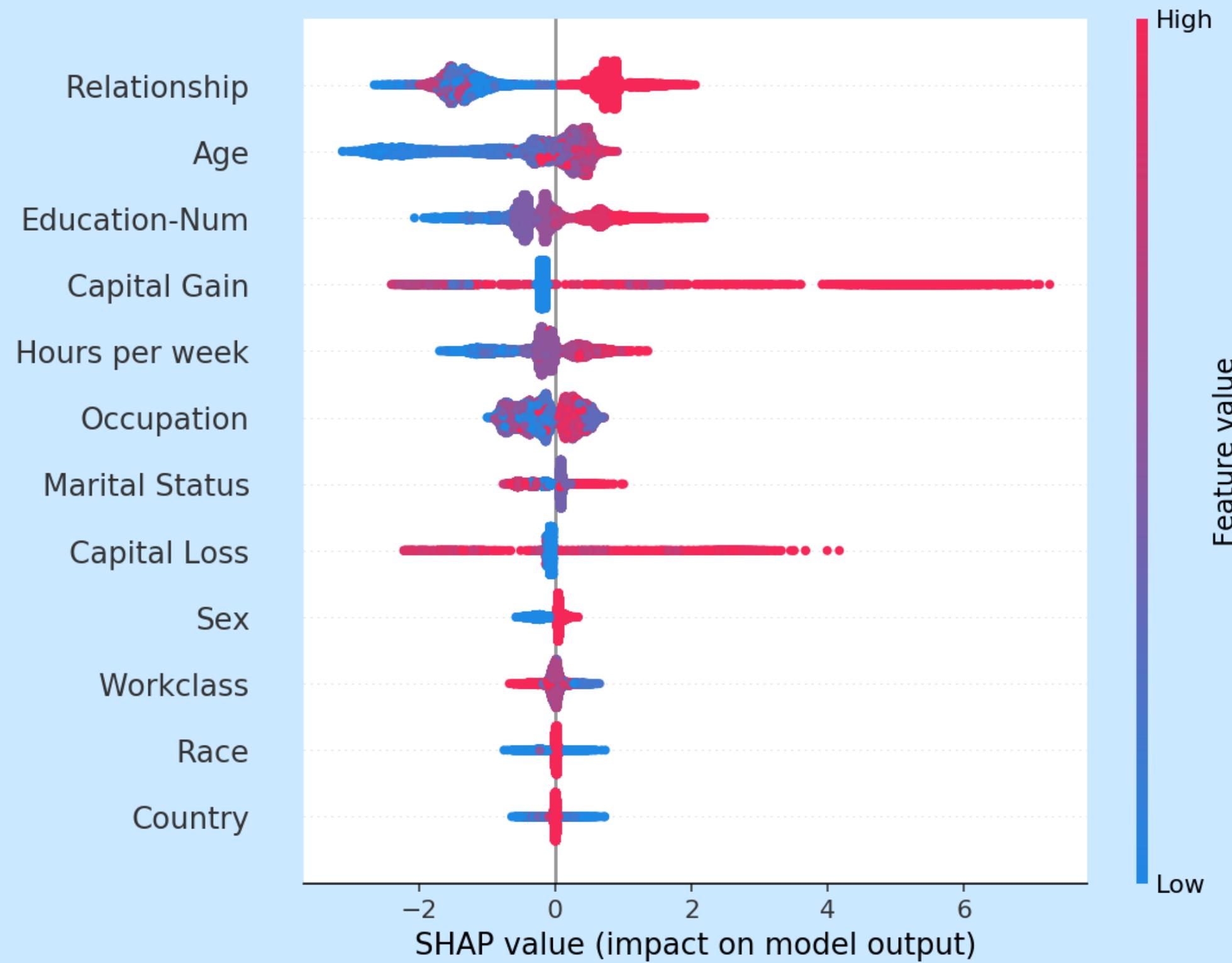
What is the SHAP method?

- ✓ a way of explaining individual predictions (and in summary)
- ✓ is based on the game theoretically optimal Shapley values (and it is quite related to LIME as well)
- ✓ The Shapley value is the average marginal contribution of a feature value across all possible coalitions.

SHAP feature importance



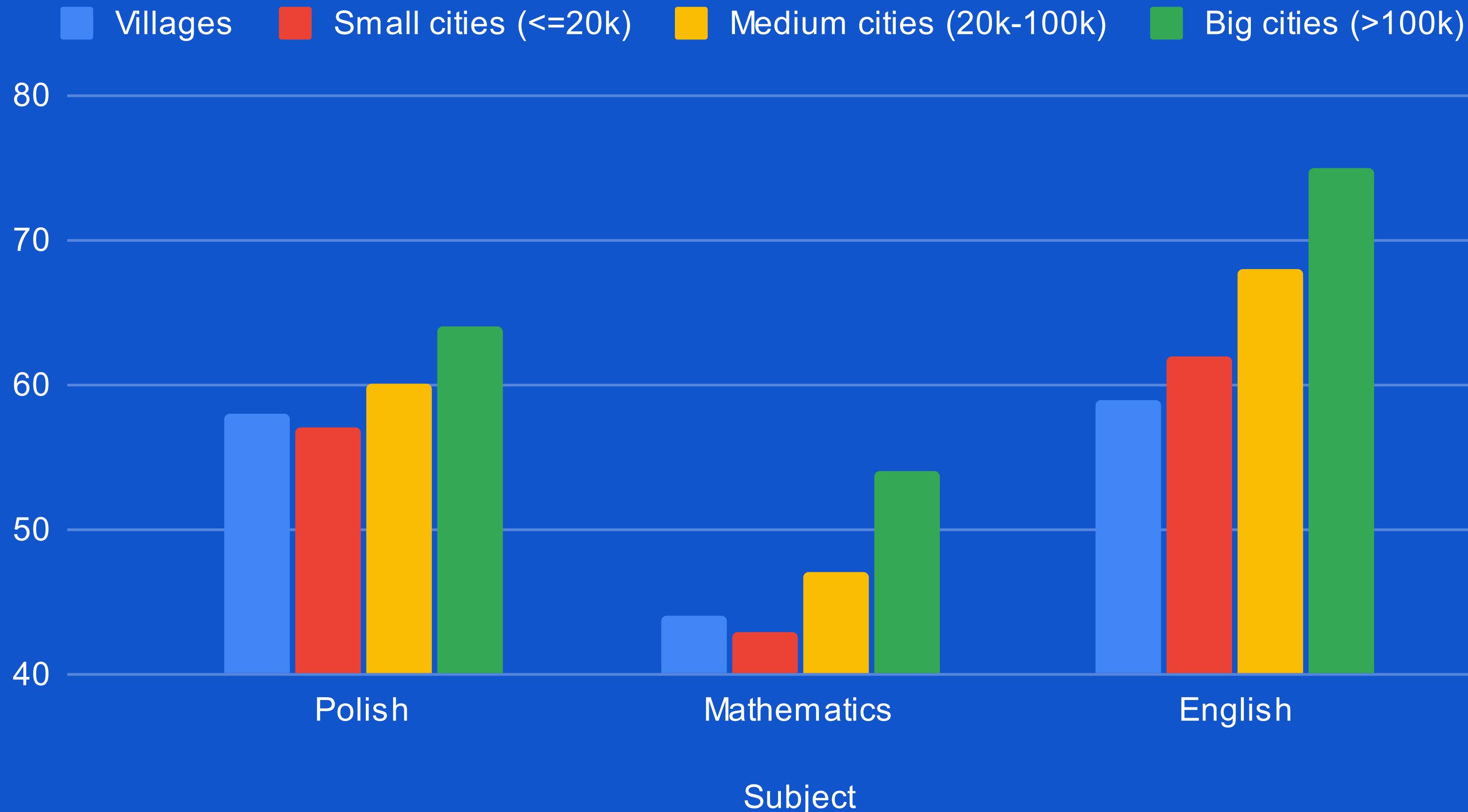
SHAP feature impact



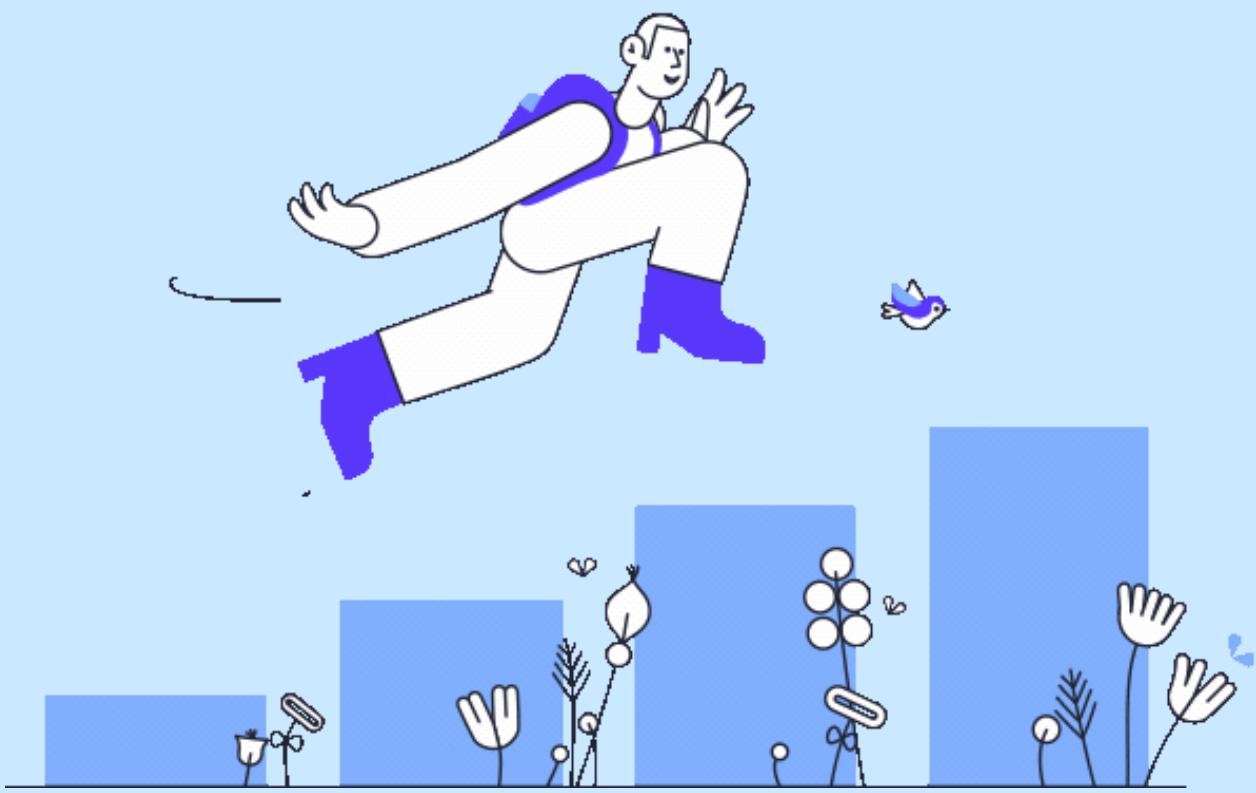


Statistics in Poland

Avg. exam scores of 8th graders in Poland by school location



"Cities overtake villages. In English"



1. Fundacja im. Stefana Batorego. (2019). *Ukryta prywatyzacja w polskiej edukacji – skala i zagrożenia*.
2. Hertel, R. F., Groh – Samberg, O. (2019) The Relation between Inequality and Intergenerational Class Mobility in 39 Countries
3. Kosz, J. (2009) *Współdziałanie w szkole – oczekiwania i potrzeba współczesnej praktyki edukacyjnej*. Forum Dydaktyczne: przeszłość, teraźniejszość, przyszłość, 2009, Numer 5/6
4. Our Kids. (2019). *Raport o szkolnictwie niepublicznym w Polsce*.
5. Sadura, P. (2017) *Państwo, szkoła, klasy*. Warszawa: Krytyka Polityczna
6. Tracz, M., (2013) *Rola edukacji w budowaniu kapitału społecznego w Polsce*. Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego, 23, 134–144.
7. UNICEF. (2018). *Niesprawiedliwy Start Nierówności edukacyjne wśród dzieci w krajach wysokorozwiniętych*.
8. Zawistowska, A. (2012). *Horyzontalne nierówności edukacyjne we współczesnej Polsce*. Warszawa: Wydawnictwo Naukowa Scholar.

Socio-economic factors



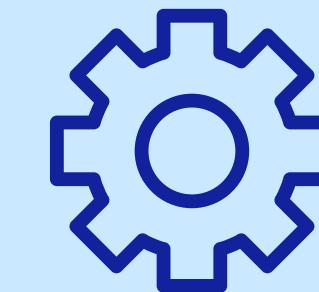
Local government revenue



School location
(village vs city)



Staff versatility

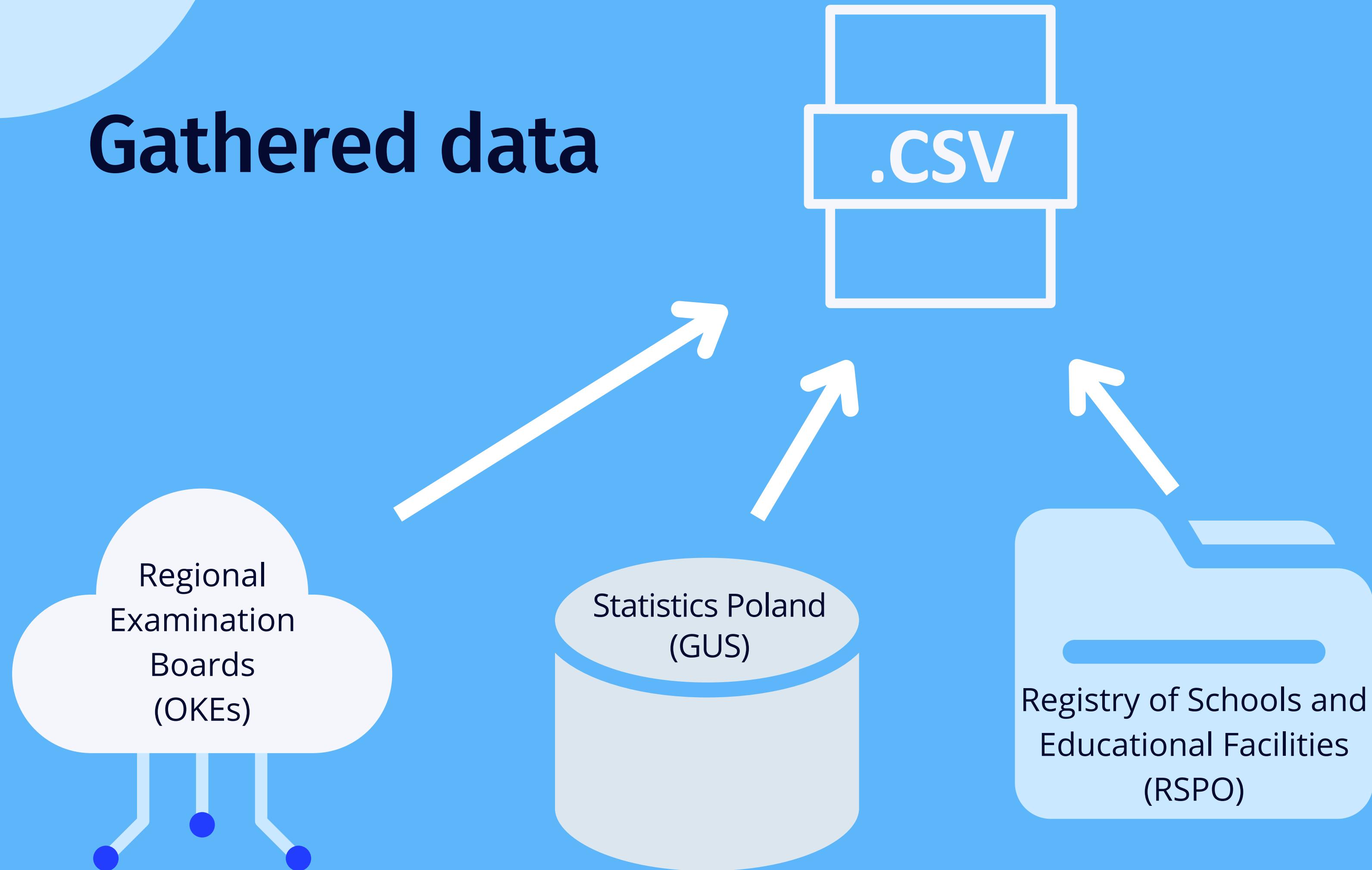


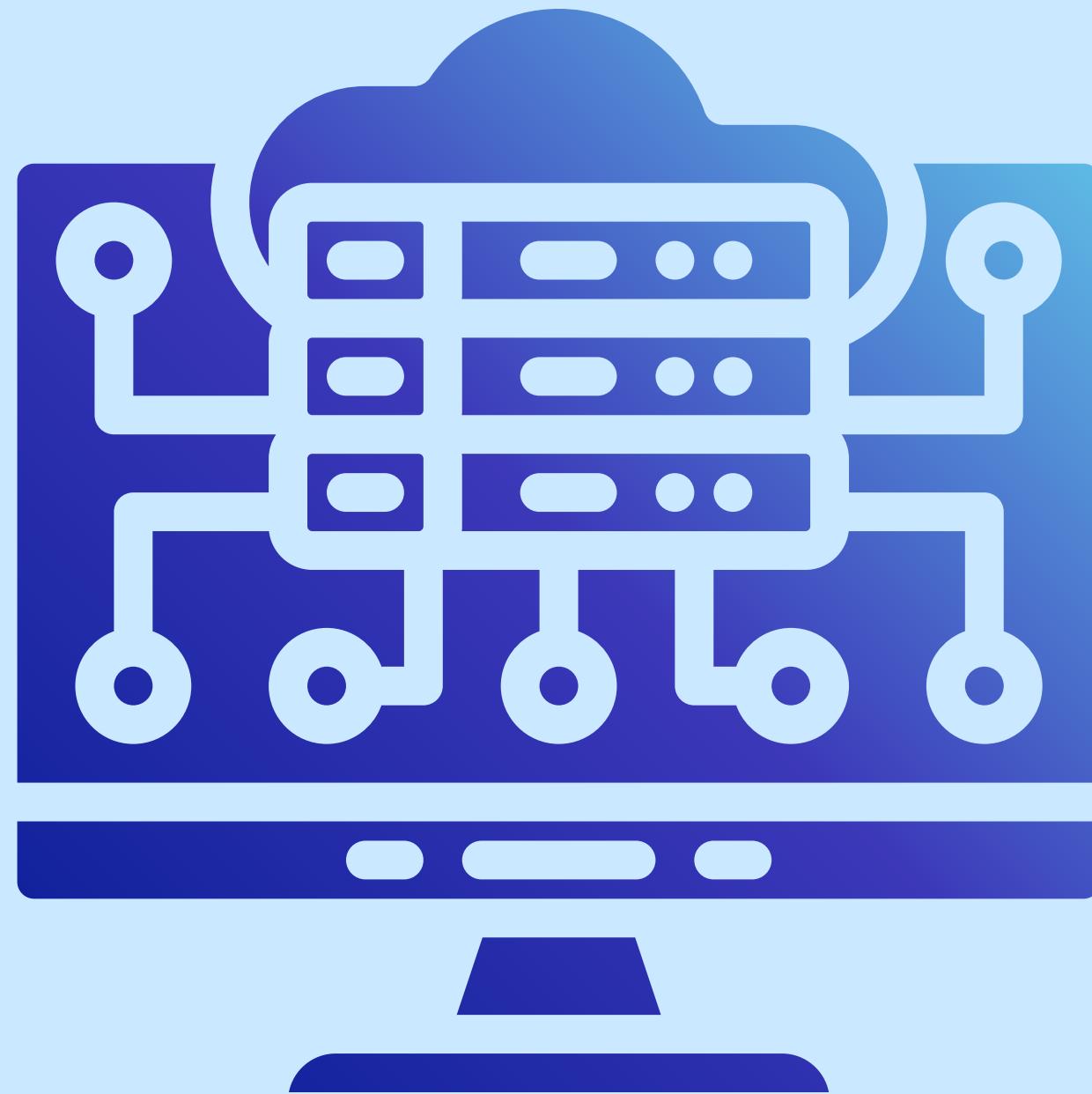
Material resources



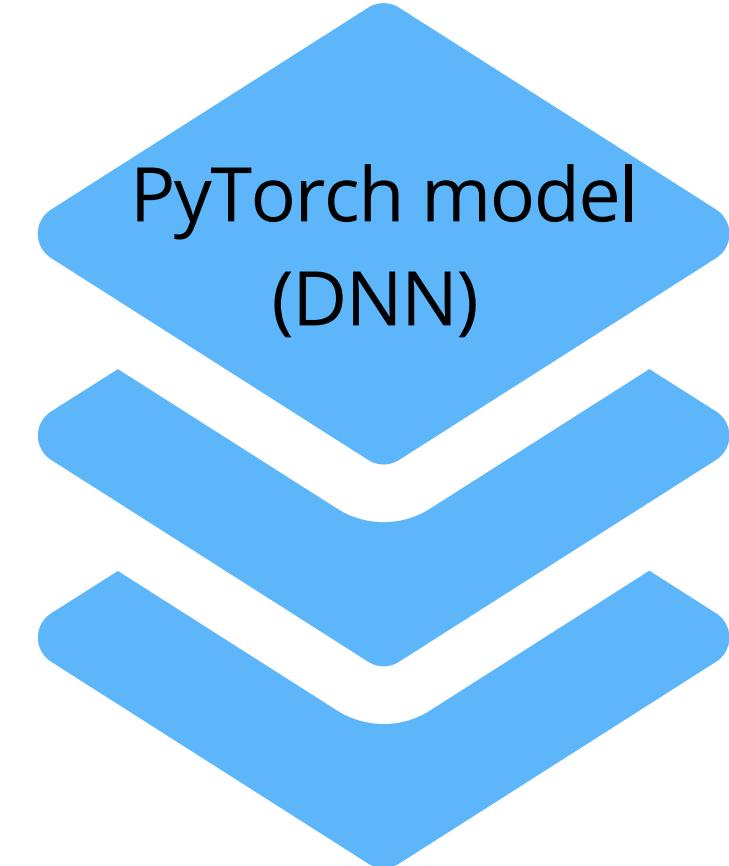
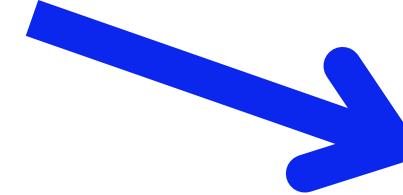
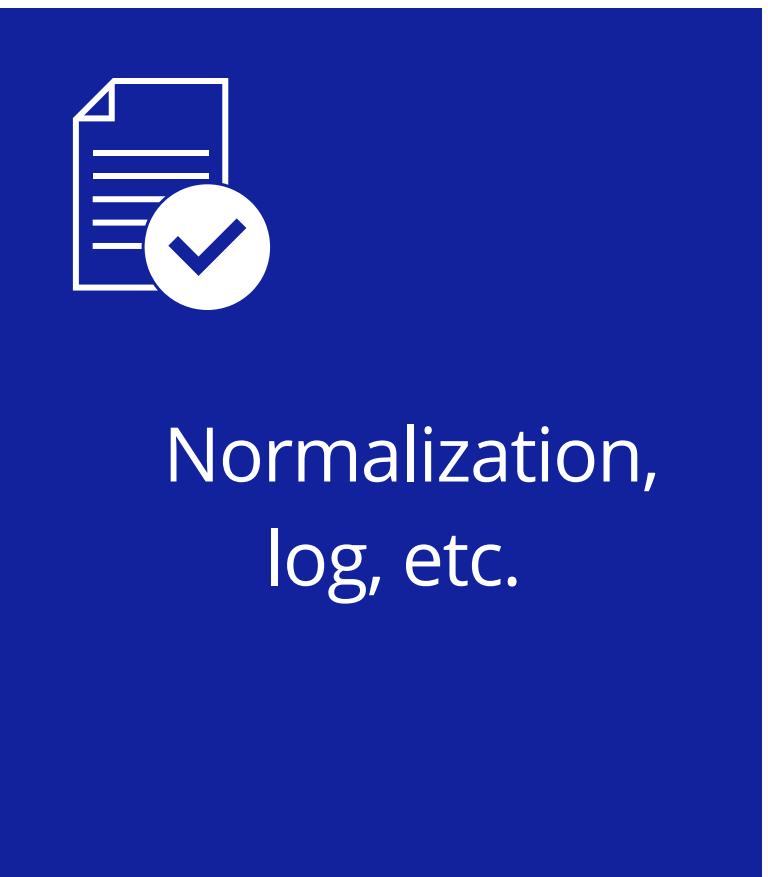
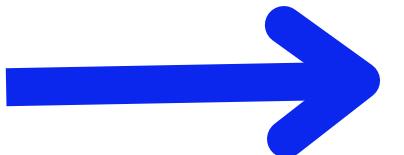
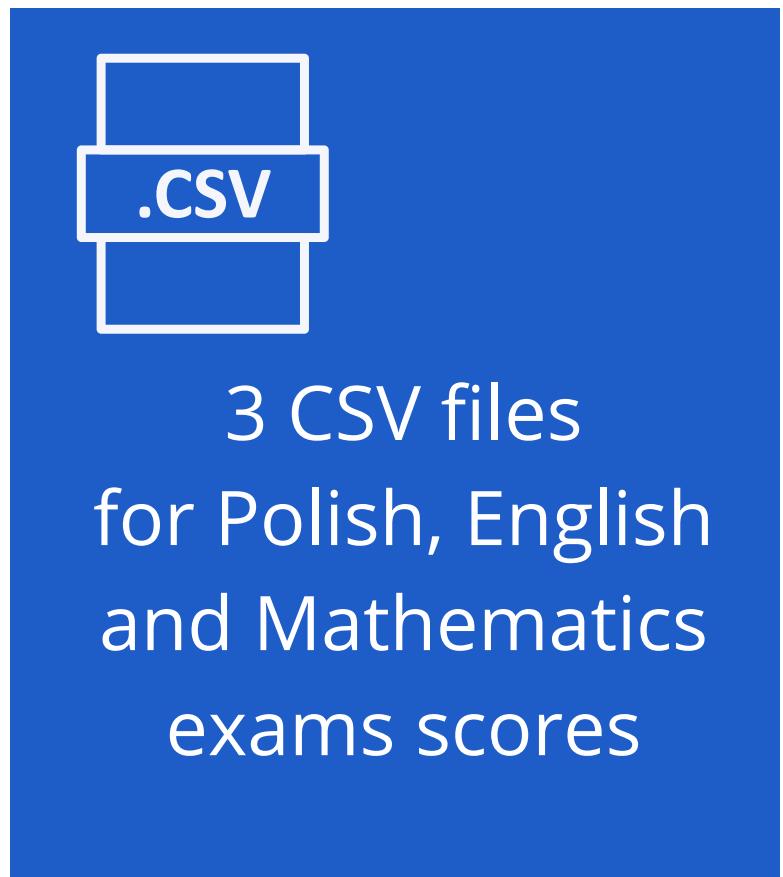
School size and age

Gathered data





Data Preparation

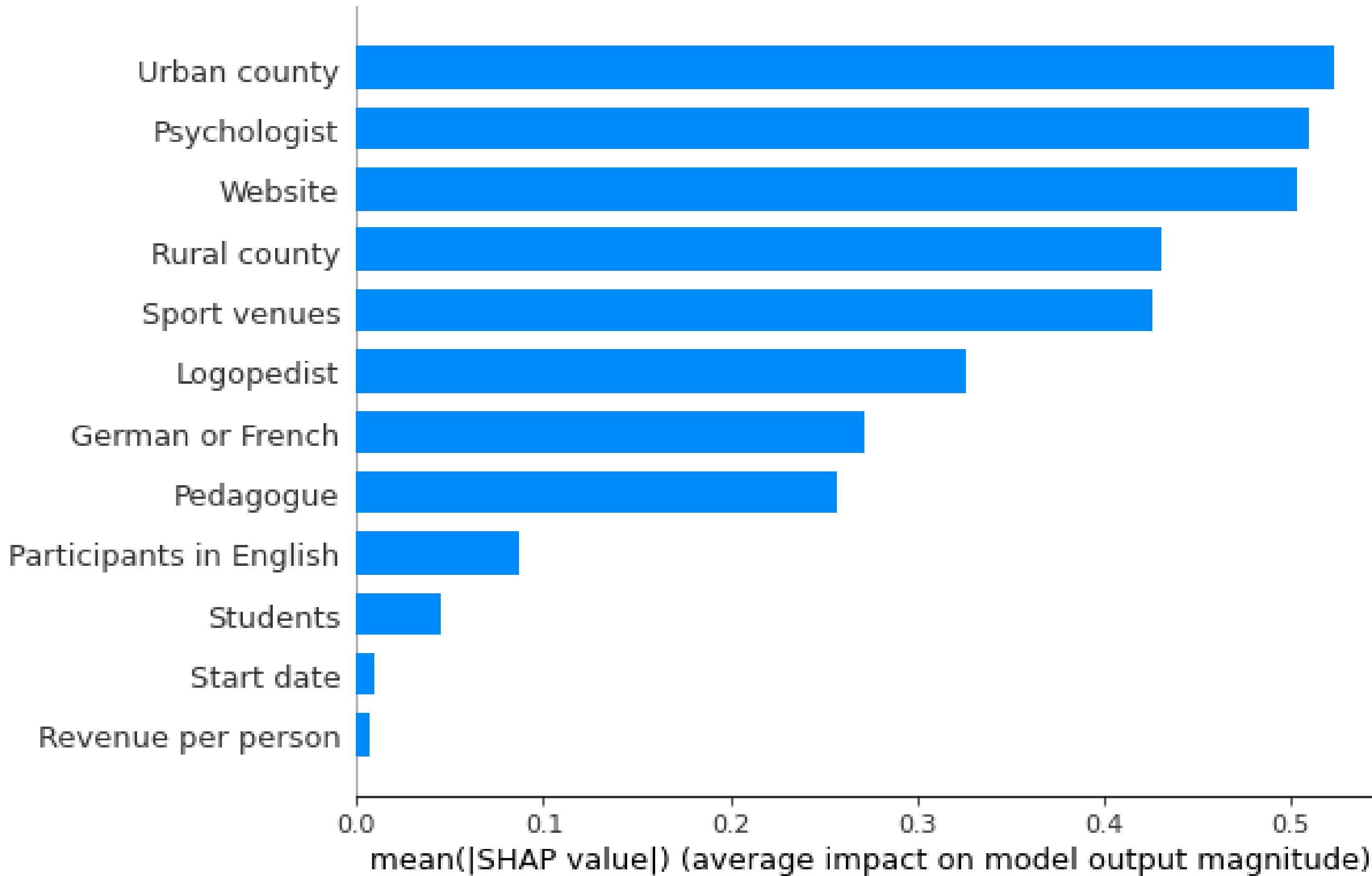


SHAP results for AutoAI (English)



**SHAP results
for Pytorch
model
(English)**

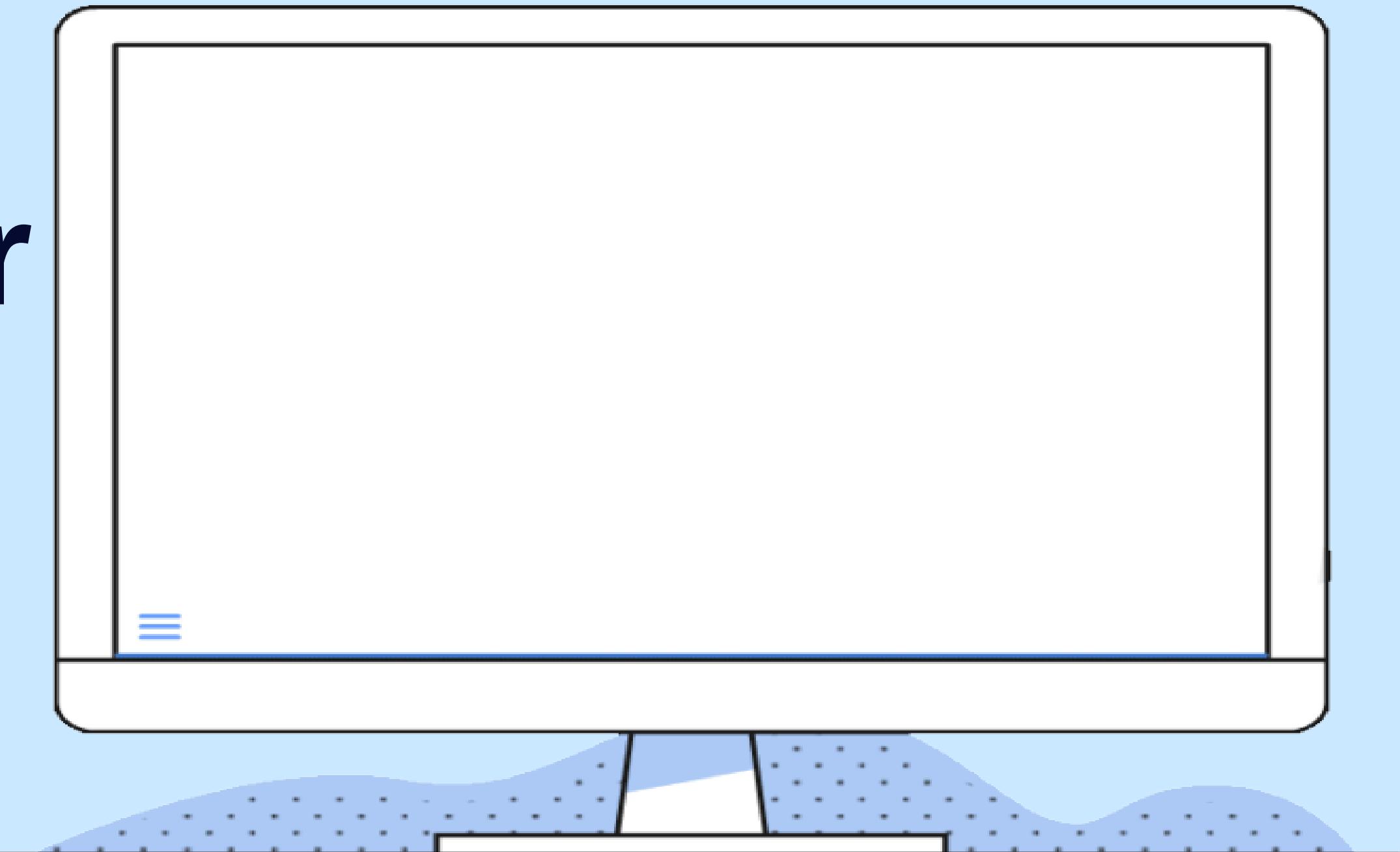




Prediction error extremes (AutoAI-English)



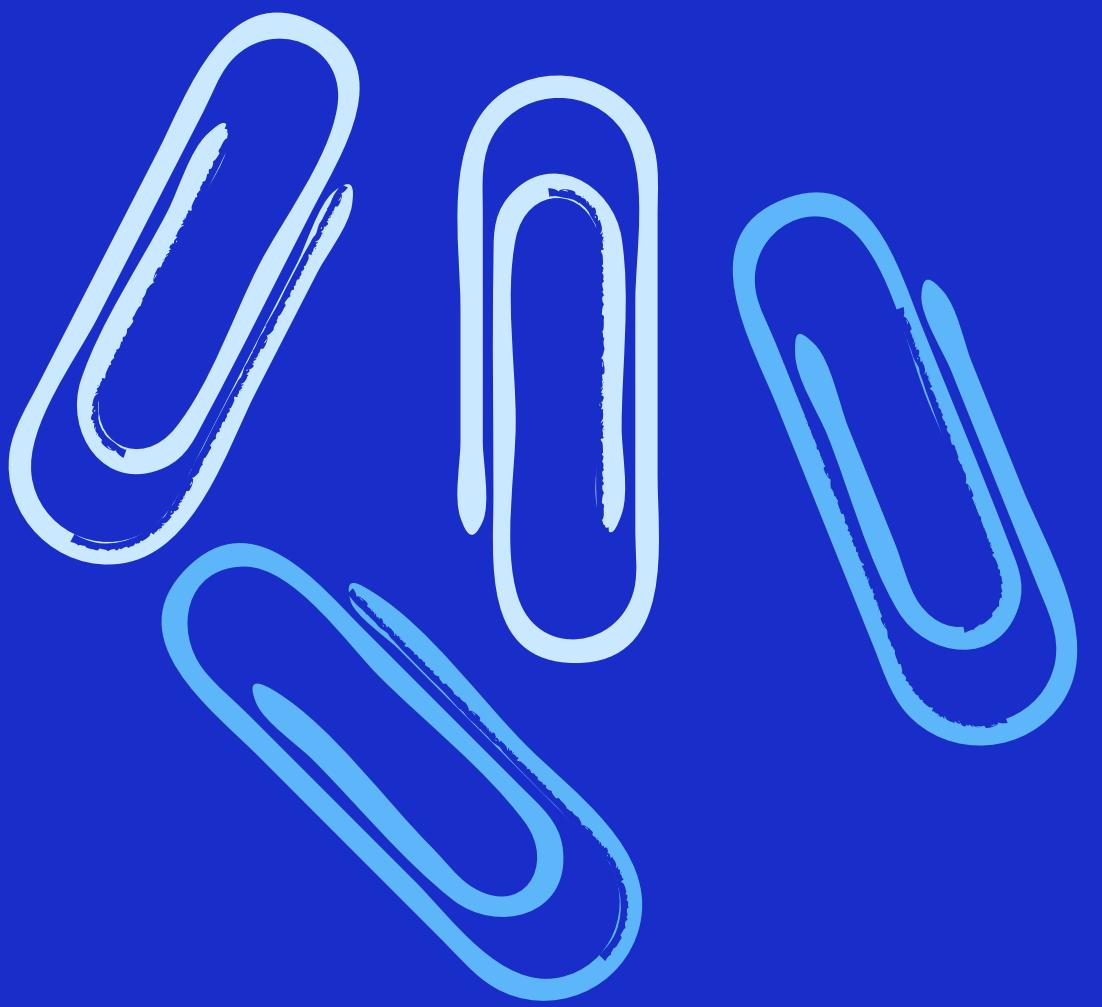
**Let's check our
website!**



References

- Census income classification with XGBoost
- 9.5 Shapley Values
- 9.6 SHAP (SHapley Additive exPlanations)
- Pexels

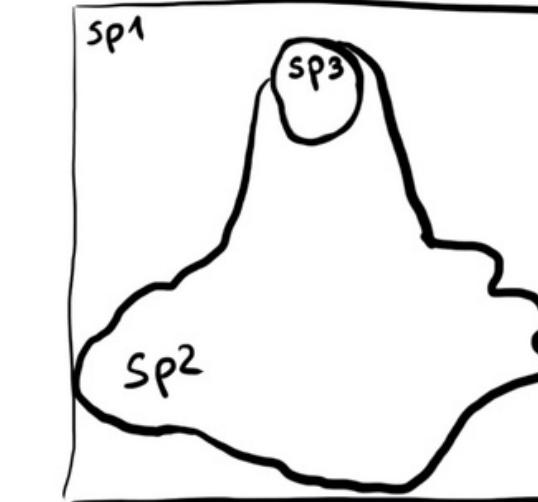
Appendix



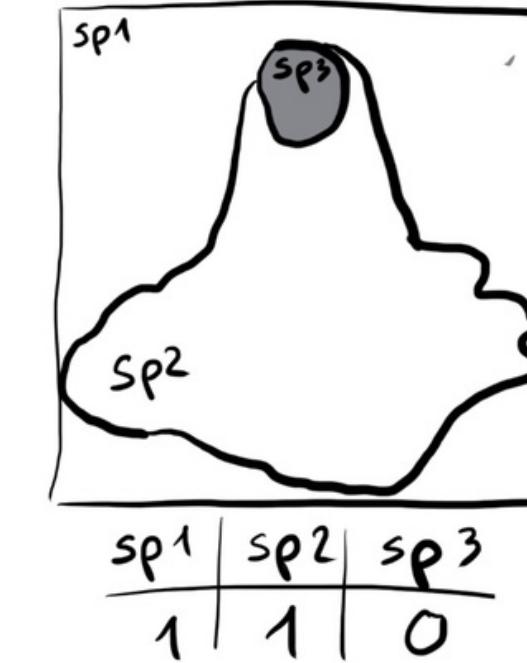
KernelSHAP vs TreeSHAP

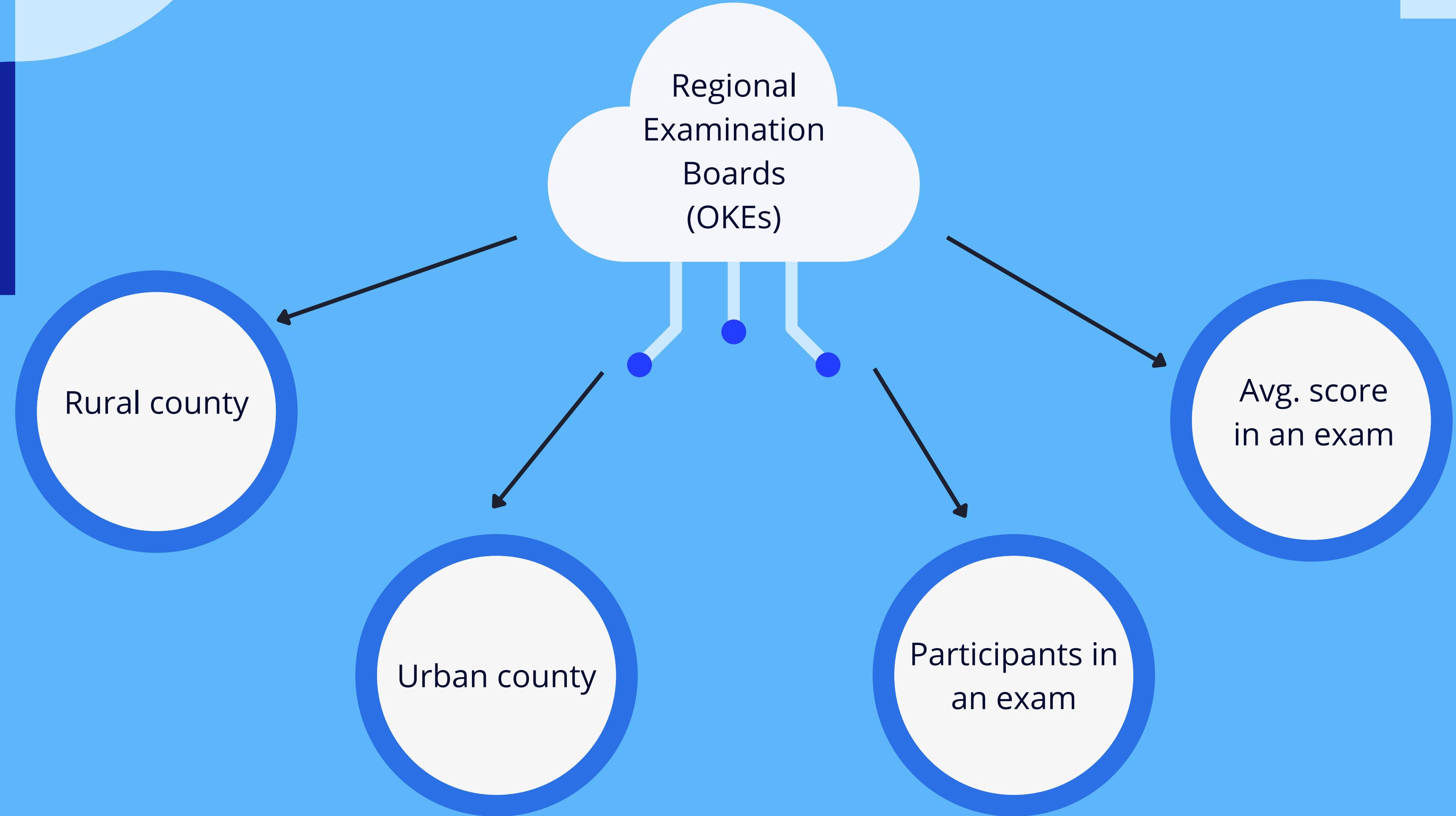
Instance x

Coalitions of super pixels $\xrightarrow{h_x(z')}$ Image



Instance x
with absent
features









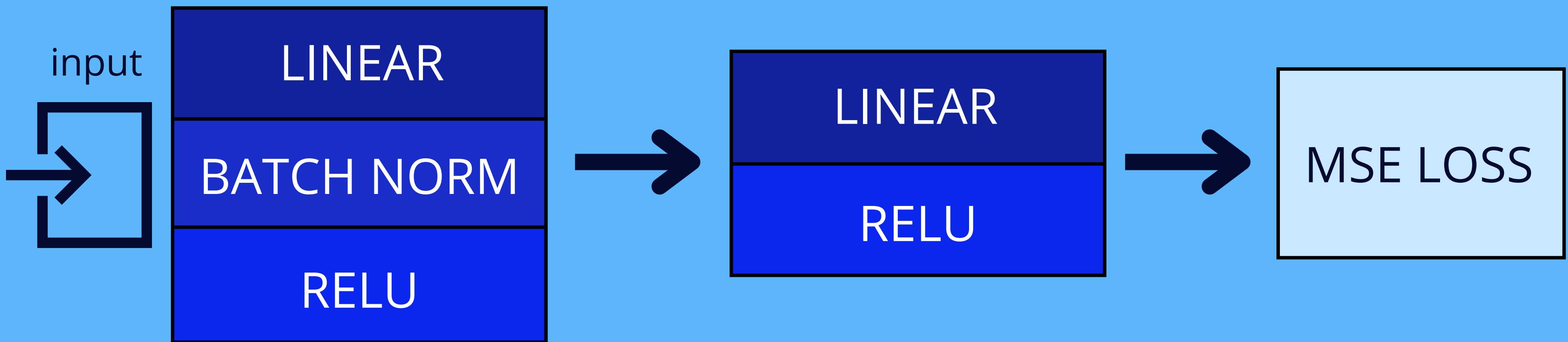
Registry of Schools and
Educational Facilities
(RSPO)



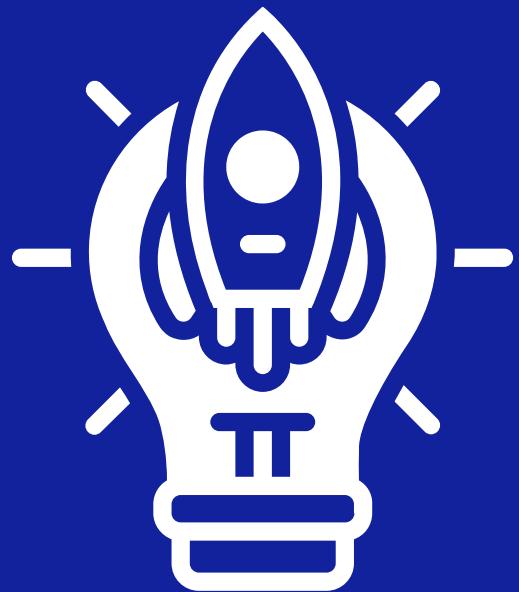
- Website
- Start date (school age)
- No. of students
- Sport venues
- Teaching German or French
- Logopedist
- Psychologist
- Pedagogue

PyTorch model architecture

3



Linear layers initialized with Xavier-normal



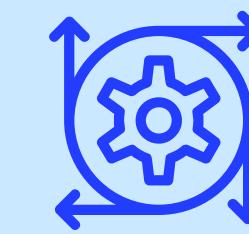
Final Thoughts



Model may base its predictions on wrong assumptions



It is essential to check what features impact the predictions and how



Data transformations may have a significant impact on the predictions and its explanations.

Thank you for your attention!

