

# PREDICTION and FACTORS of DROPOUT in HIGHER EDUCATION

## CHALLENGES

Europe in a changing world - inclusive, innovative and reflective societies

### The Industrial Problem

Every third student drops out from undergraduate programs, that is associated with social and economic costs. The shortage of skilled human resources endangers the outlook of economic growth. The goal is to predict final academic performance and identify its contributing factors.

Productive sector: Public administration

### Statistics and Mathematical Modeling Consulting Group



Offers statistical consulting and provides data science research and development service to corporate and academic partners.

### SDA Informatika Zrt.

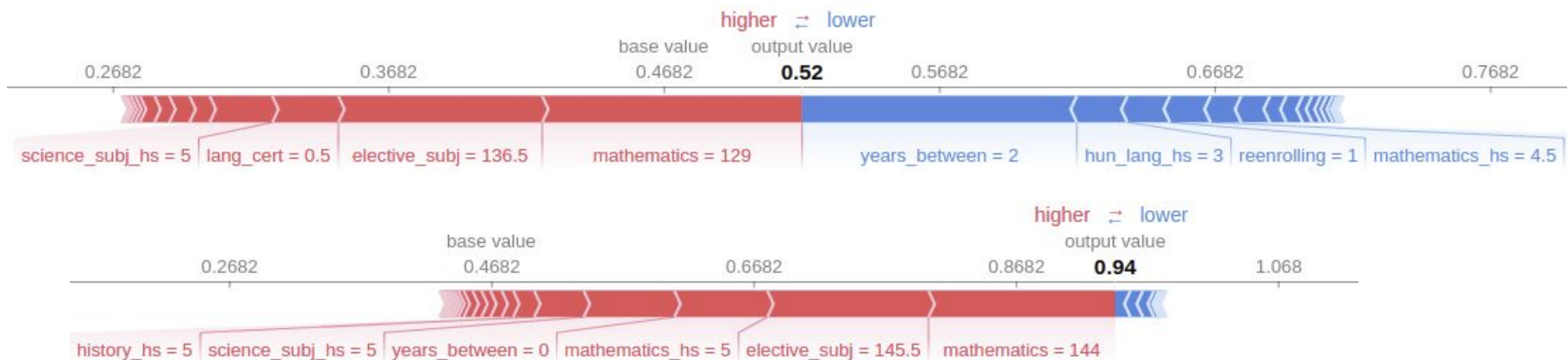


Among other things, SDA develops the educational administrative system of Hungarian universities.

# PREDICTION and FACTORS of DROPOUT in HIGHER EDUCATION

## Challenges & Goals

- To predict the **final academic performance** based on pre-enrollment achievement measures.
- To identify the **contributing factors** of university success.
- To show **how** and to what extent **features affect individual predictions** of final performance.
- To help higher education decision-makers find **at-risk students**.
- To suggest higher education policy-makers **action plans for reducing dropout rate**.
- To help students find the **skills they need to master**.

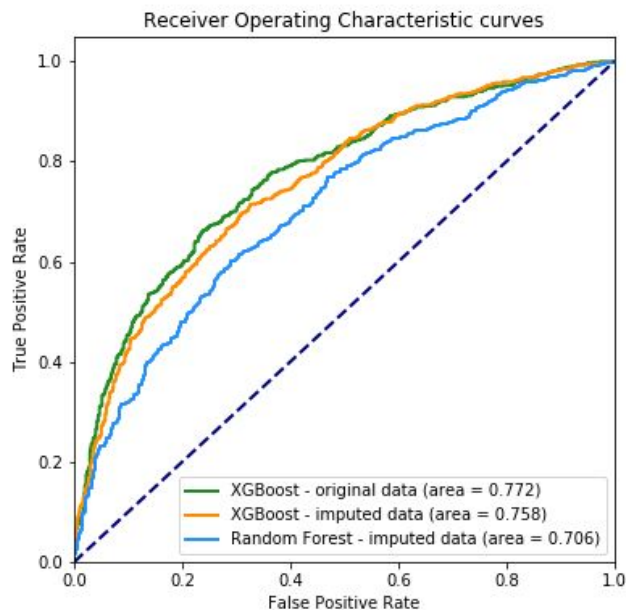


Two examples for the prediction of final performance together with the contributing factors. Features pushing the graduation prediction higher are shown in red and those pushing the prediction lower are in blue.

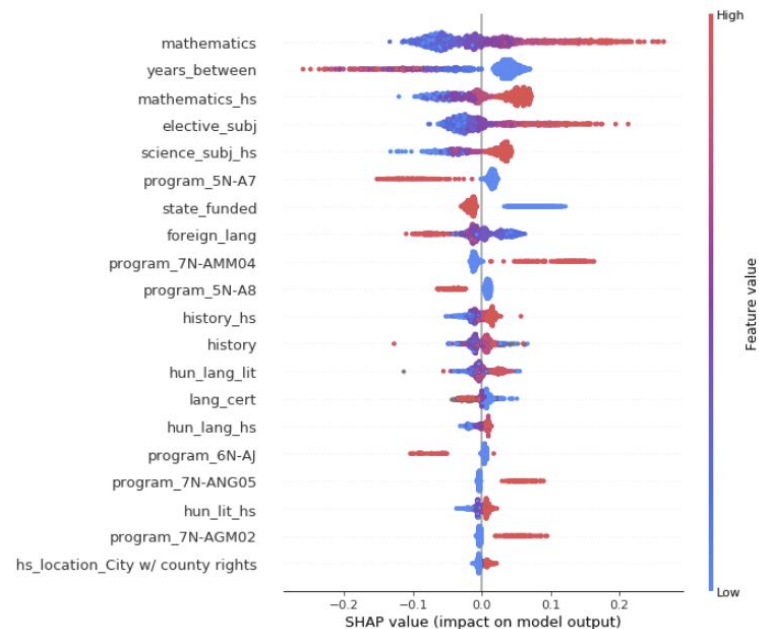
# PREDICTION and FACTORS of DROPOUT in HIGHER EDUCATION

## Mathematical and computational methods and techniques applied

- Data imputation using multiple imputation by chained equations with Bayesian ridge regression.
- Prediction of final academic performance using gradient boosted trees (XGBoost).
- Evaluation methods with receiver operating characteristic curve (ROC) analysis.
- Local model explanation with SHAP values, based on a game-theoretical concept: Shapley value.
- Global model interpretation with permutation importance and aggregated SHAP values.



*ROC curves of Xboost and Random Forest models*



*Summary of the effects on the final performance of the most important features.*

# PREDICTION and FACTORS of DROPOUT in HIGHER EDUCATION

## Results & Benefits to the company

- A **developed web application** that returns predictions of final academic performance based on some pre-enrollment achievement features. The application also highlights which skills should be improved for a given student.
- **3 years of** successful collaboration on educational research and decision support system development.
- Staff and researchers formation: **more than 10 publications in prestigious conferences and high impact journals.**

### Secondary school subjects

	Final mark (penultimate year of study)	Final mark (last year of study)
Hungarian literature	Select... ▼	Select... ▼
Hungarian language	Select... ▼	Select... ▼
History	Select... ▼	Select... ▼
Mathematics	Select... ▼	Select... ▼
Foreign language	Select... ▼	Select... ▼
Science subject	Select... ▼	Select... ▼

*User input of the web application*

The company is currently deploying our web application into the educational administrative system.