RESEARCH ON INTERPRETABILITY OF STACKING ENSEMBLE MODEL

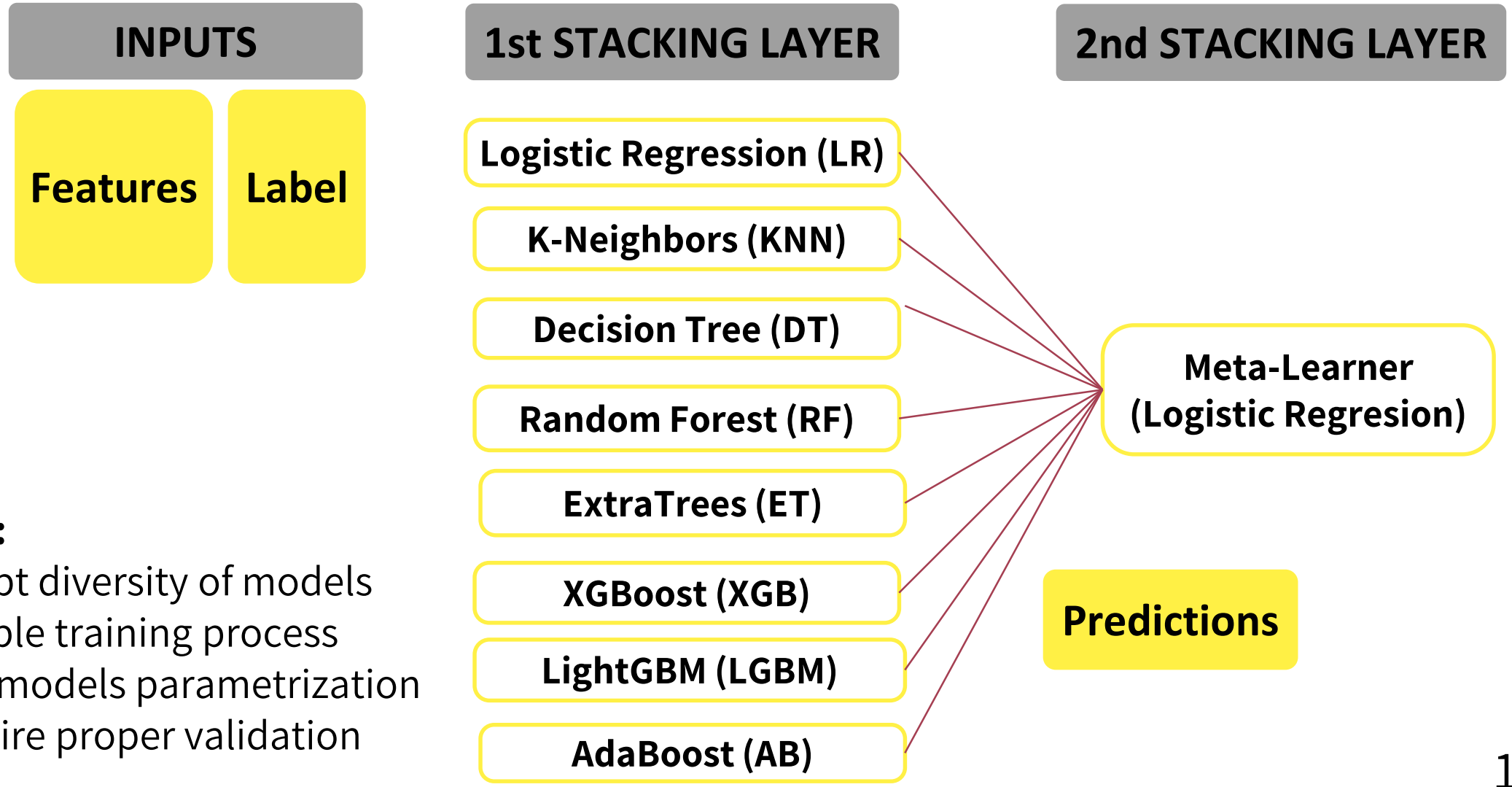
Applied to Mental Illnesses Prediction

ÁLVARO ORGAZ EXPÓSITO <alvarooe@kth.se>
HEEJE LEE <heeje@kth.se>

28/01/2019

WHAT IS STACKING?

This is the final stacking model structure of this project



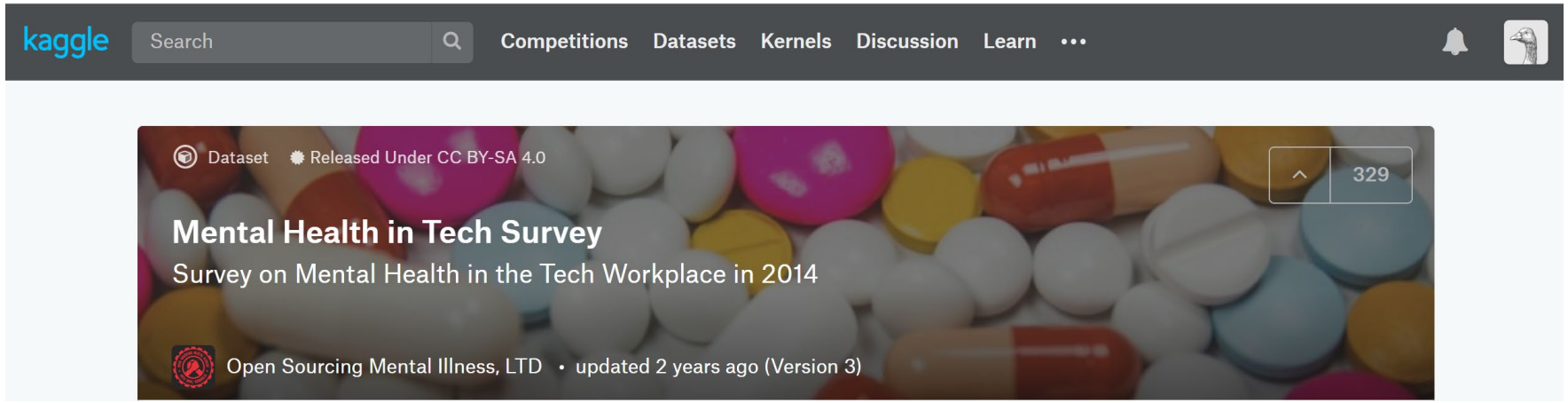
Details:

- Accept diversity of models
- Flexible training process
- Best models parametrization
- Require proper validation

WHY MENTAL ILLNESSES PREDICTION?

OSMI Mental Health in Tech Survey

Aim to measure attitudes to mental health in the tech workplace and examine the frequency of mental disorders among tech workers.



INPUTS

22 selected questions as model features

Binary label

Have you ever sought treatment for a mental health issue from a mental health professional?

PROBLEM STATEMENT

Stacking enhances predictive performance but tends to be a black-box model. Research question: Can the interpretability of stacking models predictions be improved in spite of submodels non-transparency?

Problem context

Digital era where statistics, machine learning and artificial intelligence play an important role.

Why is important to solve?

Stacking boosts predictive capacity of Machine Learning but sometimes organizations cannot use this complex technique for the lack of interpretability.

PROBLEM STATEMENT

Stacking enhances predictive performance but tends to be a black-box model. Research question: Can the interpretability of stacking models predictions be improved in spite of submodels non-transparency?

Goals

1. At population level: create indicators of feature importance, submodels importance and instability to noise.

2. At instance level: improve prediction interpretability and marginal feature effects.

Hypothesis

1. Trade-off between interpretability and performance is difficult to beat, but we can do it using statistical techniques as evaluation method.

2. We can enhance interpretability without worrying if submodels are transparent or not.

BACKGROUND AND RELATED WORK

Ensembling

Includes a wide range of techniques: averaging or blending, weighted averaging, conditional averaging, bagging, boosting, stacking, etc.

Stacking is relatively recent and there is not many documentation about it. For example, an important contribution is the thesis *StackNet* by Kaggle Competitions Grandmaster Marios Michailidis.

Feature importance studies

The feature importance was firstly introduced for random forests by Breiman.
Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

After that, there has been developed a model-independent version of the feature importance based on error rate of permuted features.
Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. "Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class"

METHOD USED TO SOLVE THE PROBLEM (1)

Empirical research: creating a statistical evaluation method

At population level
The aim is to get a better understanding of the whole stacking model

- Sub-models importance
- Feature importance
- Model instability to data noise

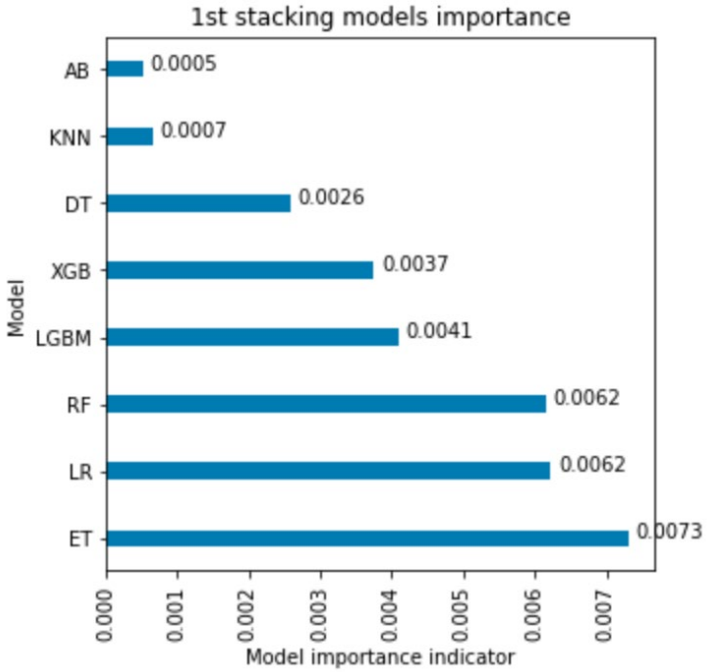
At instance level
The aim is to understand a concrete predicted value based on its features

- Accurate distribution of prediction
- Understand predictions based on features
- Marginal effects of each feature
- Feature importance

METHOD USED TO SOLVE THE PROBLEM (2)

At population level : submodels importance

Details:
- OOB sample



FOR EACH MODEL PERMUTE SEVERAL TIMES

- Logistic Regression (LR)
- K-Neighbors (KNN)
- Decision Tree (DT)
- Random Forest (RF)
- ExtraTrees (ET)
- XGBoost (XGB)
- LightGBM (LGBM)
- AdaBoost (AB)

Permuted predictions

Real predictions

Meta-Learner (Logistic Regresion)

AUC 1 | AUC 2 | AUC 3 | ...

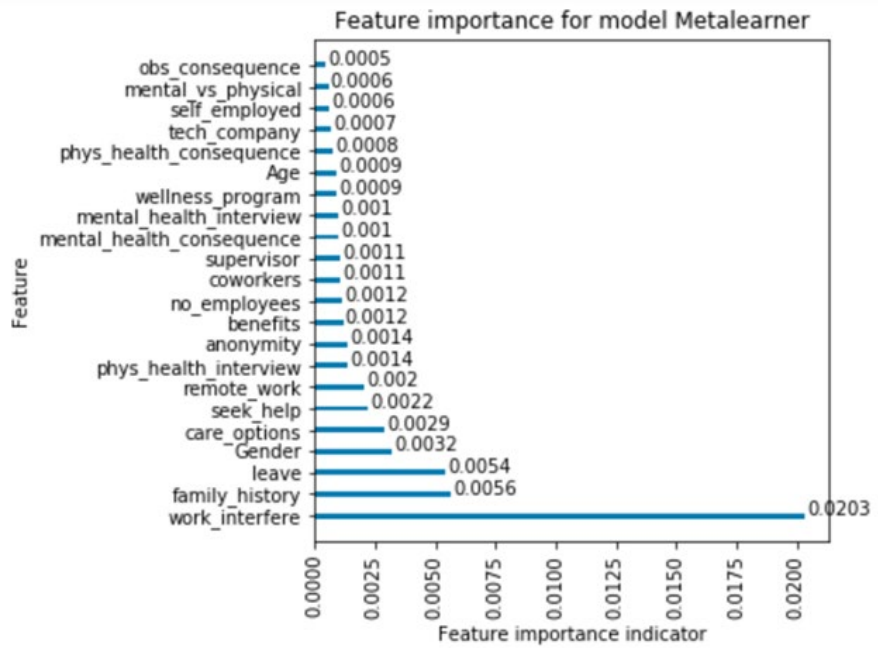
Standard Deviation

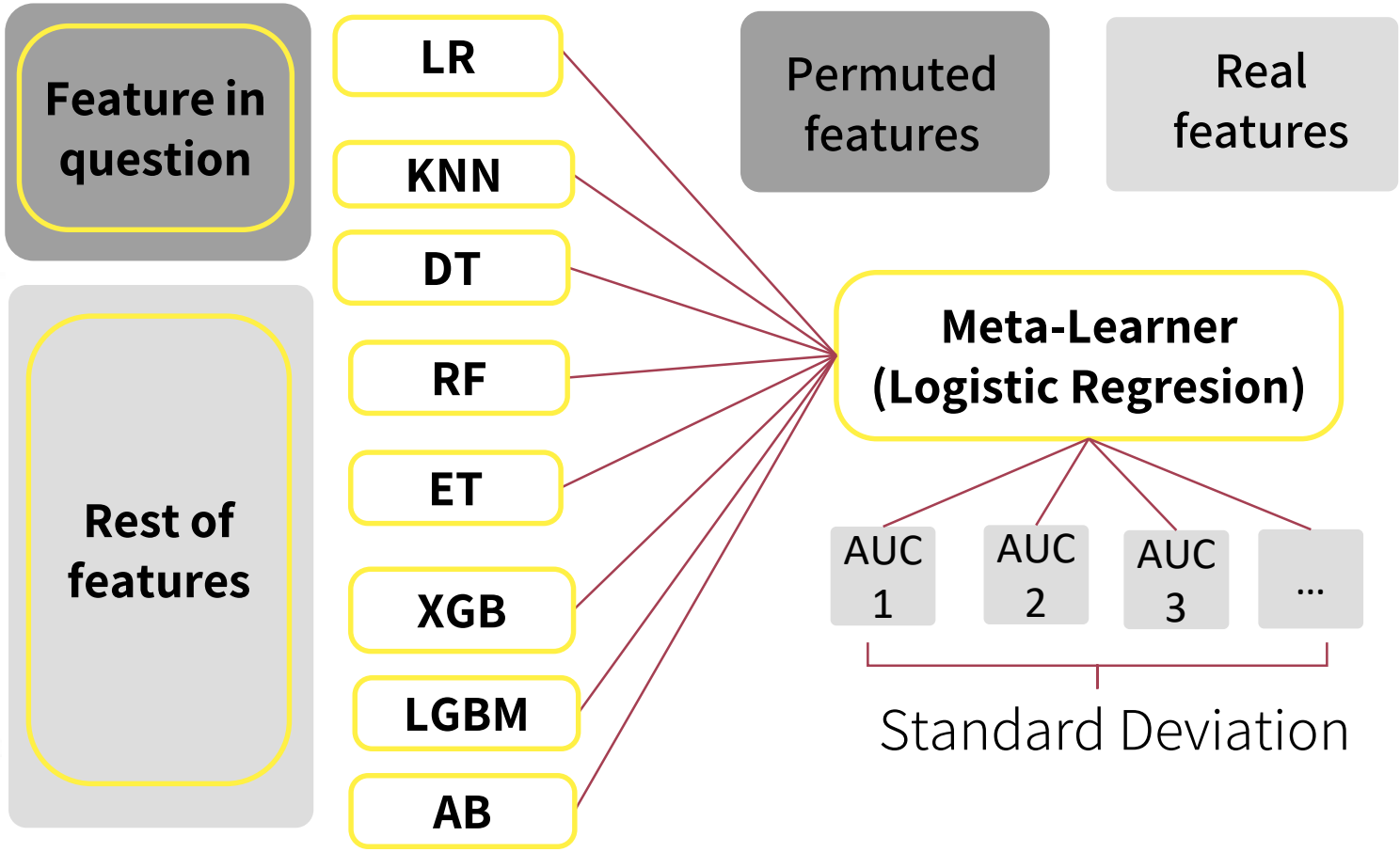
METHOD USED TO SOLVE THE PROBLEM (3)

At population level : feature importance

FOR EACH FEATURE PERMUTE SEVERAL TIMES

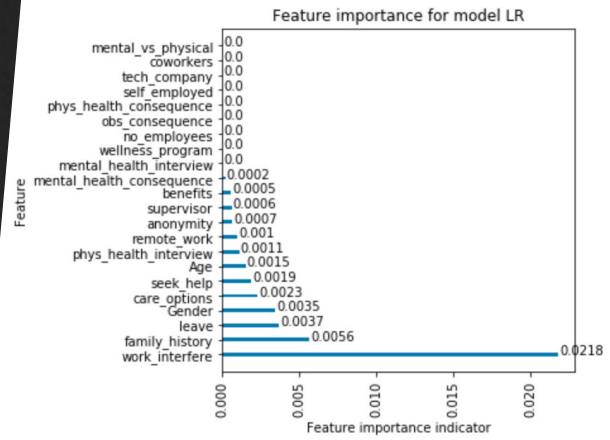
Details:
- OOB sample
- Also by submodels



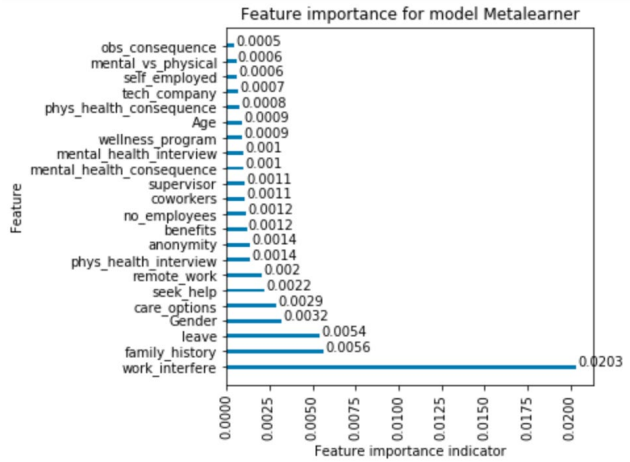


METHOD USED TO SOLVE THE PROBLEM (4)

At population level : model instability to data noise

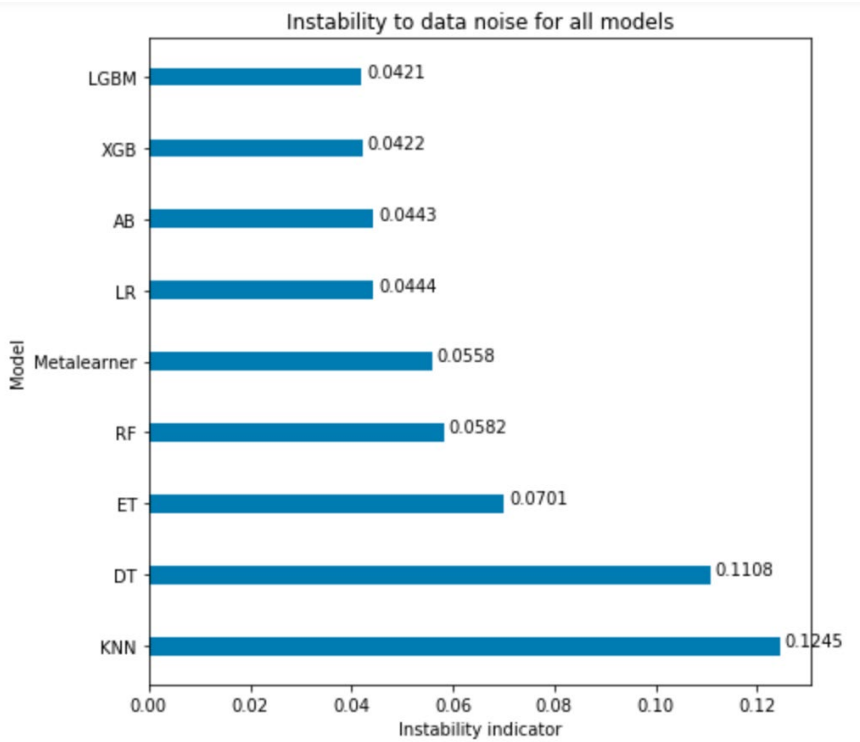


...



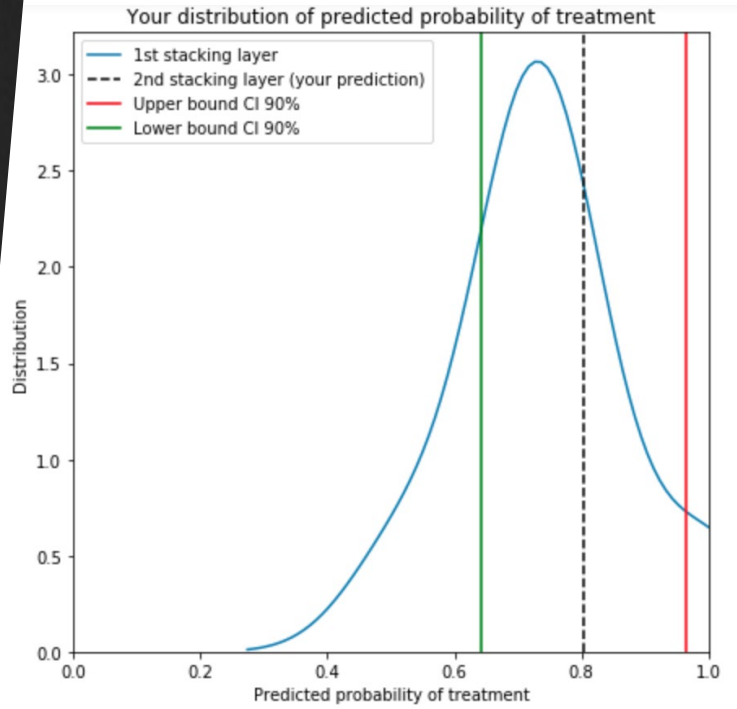
FOR EACH SUBMODEL AND META-LEARNER

Sum of all feature importances (standard deviations)



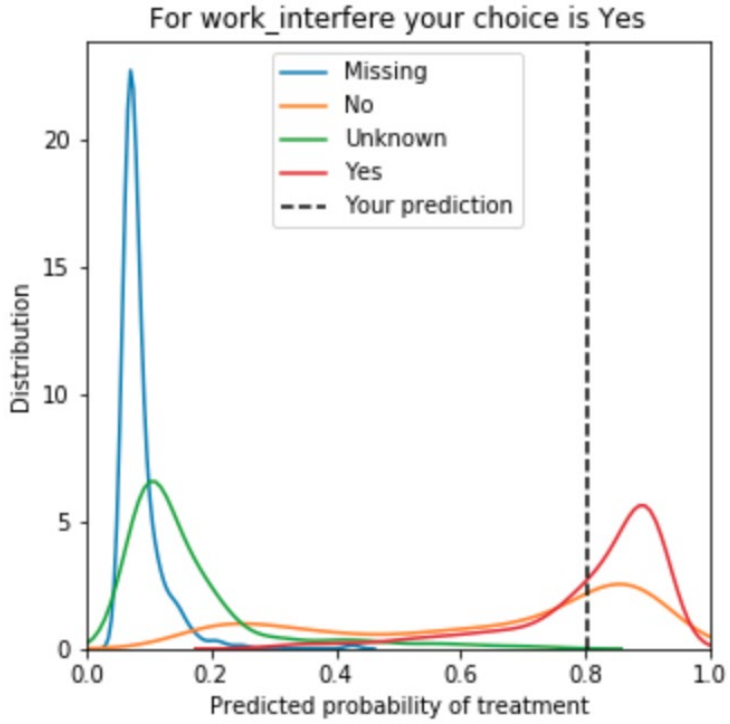
METHOD USED TO SOLVE THE PROBLEM (5)

At instance level

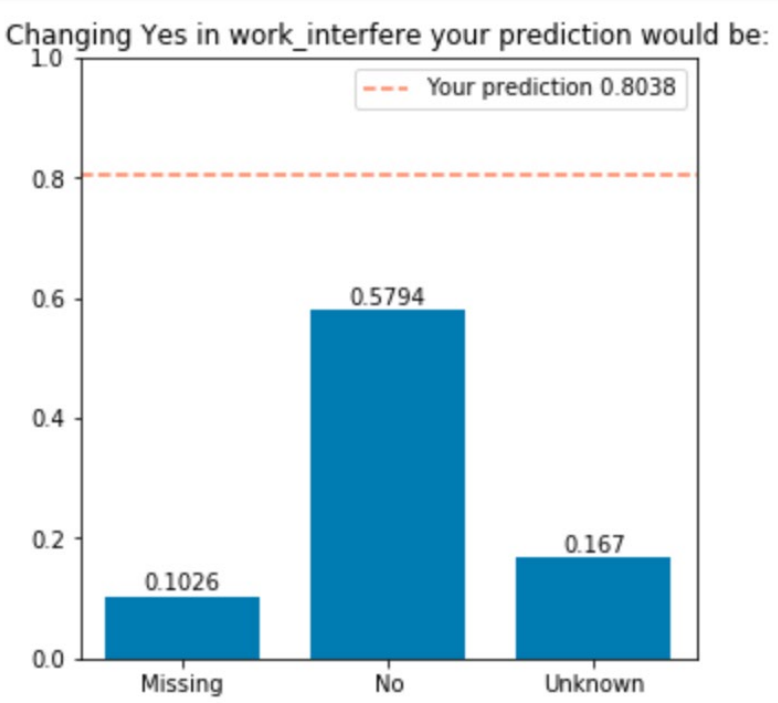


Distribution of prediction

Prediction $\pm$ Z value $*$ STD(submodels)



Understand prediction based on features values



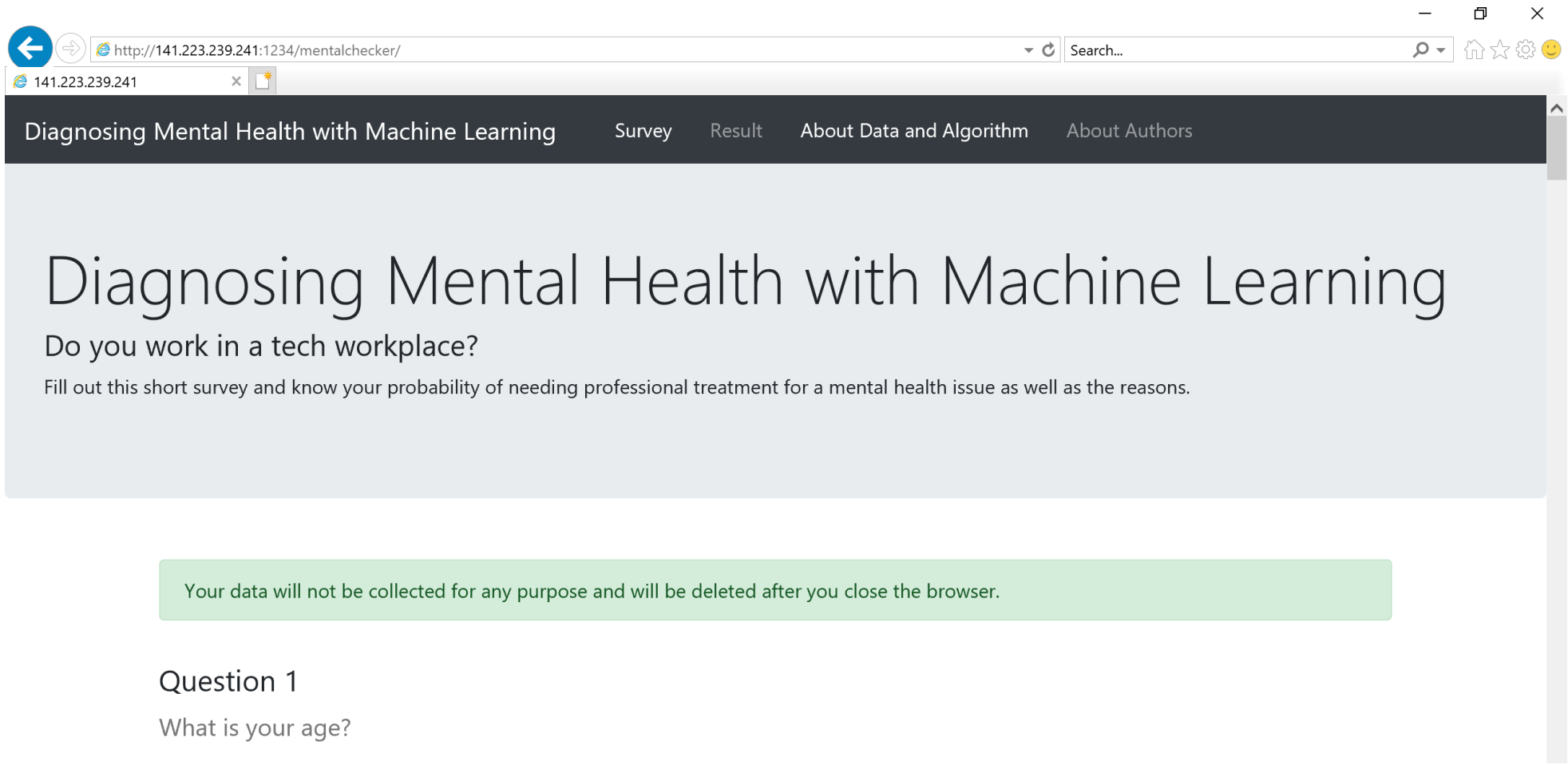
Marginal effects of each feature

↓

Feature importance:
Standard deviation of all possible predictions

RESULTS AND ANALYSIS

Web deployment at the following link:
http://141.223.239.241:1234/mentalchecker/



CONCLUSIONS

CONCLUSION for research question

Our evaluation method provides a set of statistical indicators about stacking architecture that does not depend on submodels transparency and is informative enough to answer the research question positively. However, the research method of this experiment is empirical and the results of this project could be biased, that is why it would be necessary to apply it to different datasets or stacking model architectures, and check if the evaluation method provides informative outputs too.

CONCLUSIONS for hypothesis

It is true that stacking enhances performance at the cost of interpretability, but we create a statistical evaluation method that helps to understand the stacking architecture predictive behaviour. We achieve it without worrying about the non-transparency of submodels.

DO YOU HAVE QUESTIONS?
THANK YOU

ÁLVARO ORGAZ EXPÓSITO <alvarooe@kth.se>
HEEJE LEE <heeje@kth.se>