

Machine Learning Workshop

May 27, 2021

ResponsibleML based models
for prioritizing liver-transplant candidates in the US

Hoang Thien Ly

Outline

1/ Transplant problem and sickest-first policy

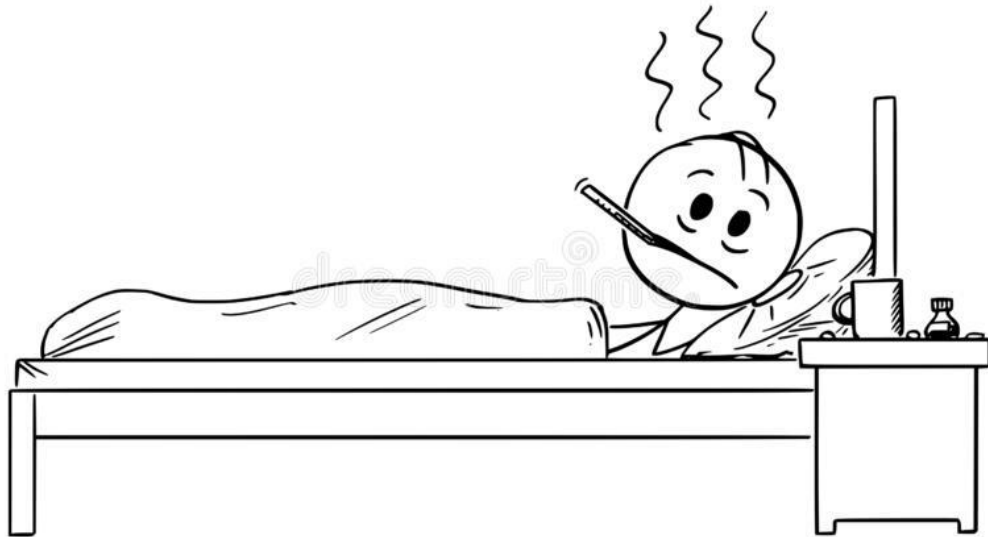
2/ MELD-Score and optimistic results

3/ Applying ML Techniques

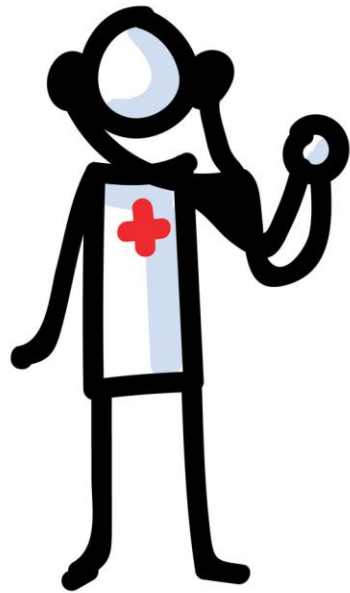
- Process of cleaning datasets
- Models and obtained results from a paper

4/ Related works and Q&A section

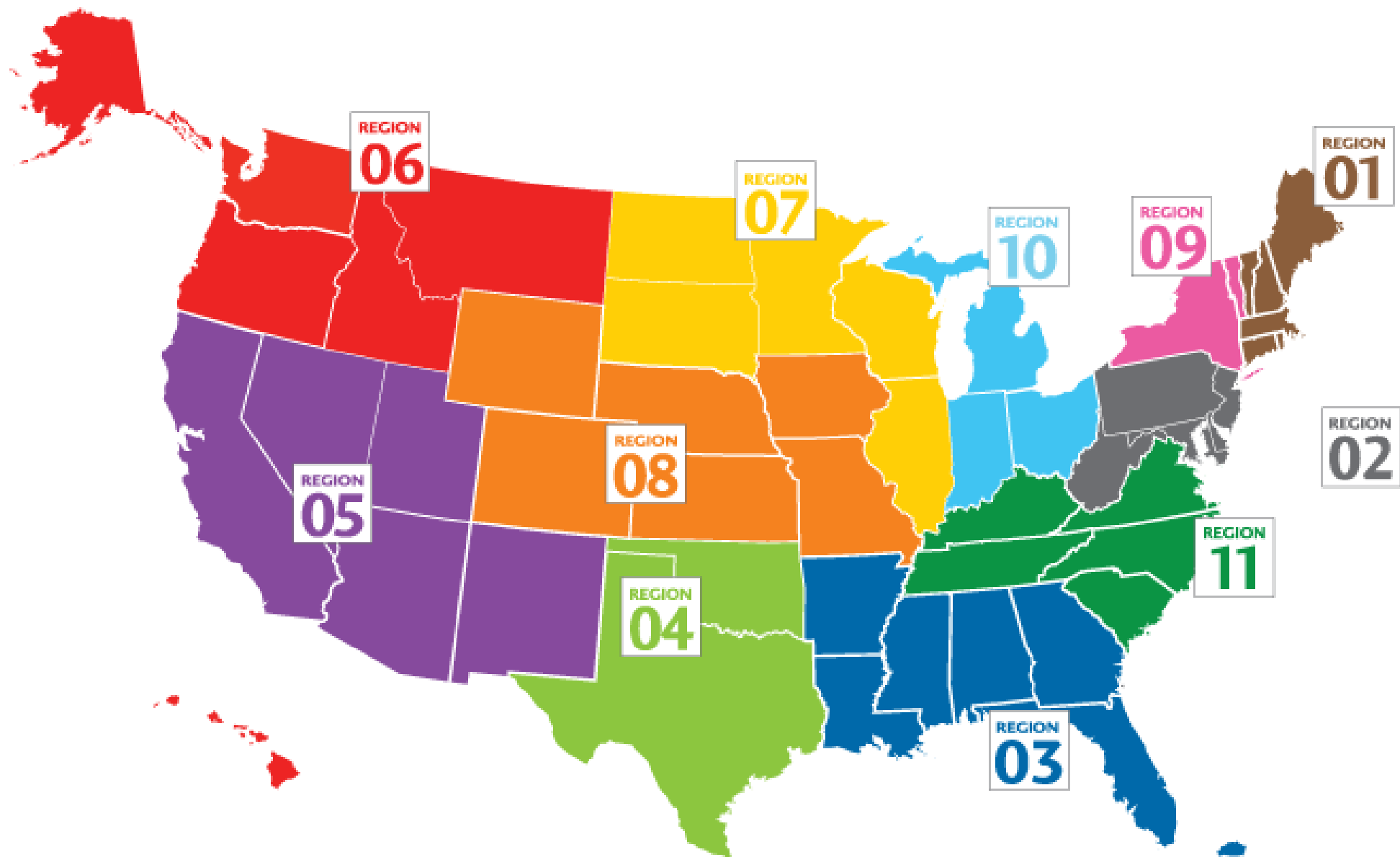
Transplant problem & policy

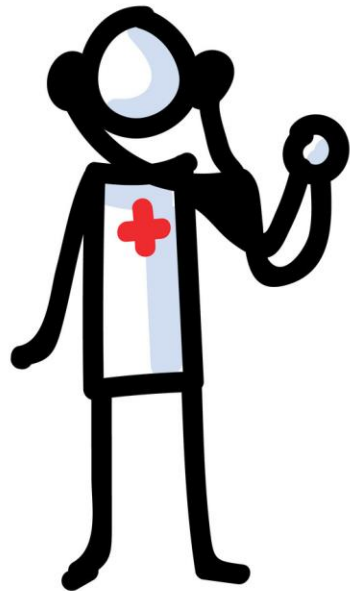


I am patient with end-stage liver disease. I need a liver-transplantation.

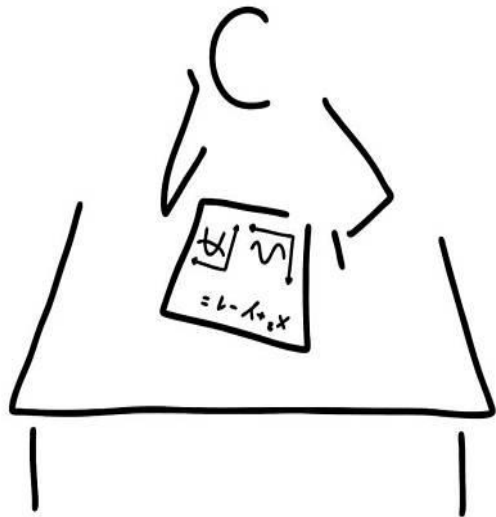


- One national waiting list for the entire US. You can register at medical centers in different regions.
- Waiting time cannot be predictable and may vary from regions of the country and from the illness levels.





- The deceased donor livers are assigned to matches among patients according to sickest-first policy.
- When get top of the waiting list, we consider compatibility of patient to donor liver to lower risk of failure:
 - + Blood type,
 - + Body size,
 - + Geographical considerations...



- November 12, 2019:
12 965 patients are on waiting list for liver transplantation.
- In 2018:
12 700 patients were added to the list.
8 250 liver transplants were performed.
609 patients died while waiting.
627 patients were removed due to being too sick.

Problem of assessing
disease severity of a patient

Outline

1/ Transplant problem and sickest-first policy

2/ MELD-Score and optimistic results

3/ Applying ML Techniques

- Process of cleaning datasets
- Models and obtained results from a paper

4/ Related works and Q&A section

MELD-score and its variants

- The **M**odel for **E**nd-stage **L**iver **D**isease (MELD-score) estimates the chance of surviving during the next 3 months of patients with chronic liver disease.
- MELD-score is ranged from 6 to over 40
- The higher the score, the sicker the patient is, and the higher his/her position is in waiting list
- It has variants: MELD, MELD-Na, MELD-exception (for liver cancer, or hepatopulmonary syndrome,...) and PELD (for pediatric patients)

MELD-score and its variants

1 Year Survival Rate Based on MELD Scores

<u>MELD Score</u>	<u>Waiting list</u>	<u>Post Transplant</u>
Score 10	90% survival	83% survival
Score 15	81% survival	80% survival
Score 20	63% survival	78% survival
Score 25	42% survival	74% survival
Score 30	21% survival	71% survival

Formula of MELD-Score

$$\text{MELD} = 3.78 \times \ln [\text{Bili (mg/dL)}] + 11.2 \times \ln [\text{INR}] + \\ + 9.57 \times \ln [\text{Creati (mg/dL)}] + 6.43$$

$$\text{MELD-Na} = \text{MELD} + 1.32 \times (137 - \text{Na}) - [0.033 \times \text{MELD} \times (137 - \text{Na})]$$

Bilirubin: how well liver clears substance “bile” (żółć)

INR: how well liver makes proteins needed for blood to clot (krzepnięcie krwi)

Creatinine: how well kidneys work

Na: serum sodium, recently added, how well body regulates fluid balance.
Ranged from 125 to 137

Optimistic results from MELD-Score

The MELD-based allocation system was immediately successful, leading to the first ever reduction in the number of waiting list candidates and a 15% reduction in mortality among those on the waiting list.

Freeman, R., Wiesner, R., Edwards, E., Harper, A., Merion, R., Wolfe, R.: Results of the first year of the new liver allocation plan. Liver Transplant, **10**, 7-15 (2004)

Drawbacks of MELD-Score based system

- The log-transformed values of Bili, INR, Creati at 1.0 can be problematic, as a large percentage waiting list candidates possess Creati levels below, and values below this threshold can reflect different levels of kidney function.

Sharma, P., Schaubel, D., Sima, C., Merion, R., Merion, R., Lok, A.: Re-weighting the model for end-stage liver disease score components. *Gastroenterology*, **135**, 1574-1581 (2008)

- Correlation between MELD and outcome is not equally strong for all patients. For some patients, MELD may not accurately reflect the severity of their condition.

It's high time for ML Models!!

Outline

1/ Transplant problem and sickest-first policy

2/ MELD-Score and optimistic results

3/ Applying ML Techniques

- Process of cleaning datasets
- Models and obtained results from a paper

4/ Related works and Q&A section

Applying ML methods

- Data: waiting list histories from January 1, 2002 to September 5, 2021
- Data division: 60-40% for train-test
- Out of sample test set.
- Metric: ROC AUC (concerning with giving livers to patients who most need them than accurately predicting mortality risk)
- Number of features: 426, ~330 000 observations
 - Some known at registration
 - Some updated over time

Data Preprocessing

- Remove columns having $> 30\%$ NAs
- Num features with NAs \leftarrow median
- Cat features \leftarrow dummy variables
- Data is quite balanced.
- Train models using two different feature sets:
 - **First set:** non demographic features: Bilirubin, INR, Creatinine, Albumin, Sodium and several lab-based dialysis in previous weeks
 - **Second set:** first set including Bmi_Calc and Gender.

Training model

- Models: randomForest, catboost, XGBoost & LightGBM by {forester}
- Detail at: <https://github.com/ModelOriented/forester>
- XAI methods: <https://github.com/ModelOriented/modelStudio>
- Checking fairness: <https://github.com/ModelOriented/fairmodels>



Obtained results

```
> modell1 <- forester(X_train1[sample(nrow(X_train1),5000), ],
There were 16 warnings (use warnings() to see them)
+           target= "Transplanted_perform",
+           data_test = X_test1[sample(nrow(X_test1), 6000), ],
+           typ = "classification", metric= "auc")
```

FORESTER

Original shape of train data frame: 5000 rows, 30 columns

NA values

There is no NA values in your data.

CREATING MODELS

```
--- Ranger model has been created ---
--- Catboost model has been created ---
--- Xgboost model has been created ---
--- LightGBM model has been created ---
```

COMPARISON

Results of compared models:

model	auc	recall	precision	f1	accuracy
Catboost	0.9973786	0.9947572	1.0000000	0.9973717	0.9961667
Ranger	0.8651581	1.0000000	0.9097885	0.9527636	0.9275000
XGboost	0.5000000	1.0000000	0.7311667	0.8447097	0.7311667
LightGBM	0.4998860	0.9997721	0.7311219	0.8445985	0.7310000

The best model based on auc metric is Catboost.

There were 14 warnings (use warnings() to see them)

Obtained results

```
> model2 <- forester(X_train2[sample(nrow(X_train2), 5000), ],  
+                   target= "Transplanted_perform",  
+                   data_test = X_test2[sample(nrow(X_test2), 6000), ],  
+                   typ = "classification", metric= "auc")
```

FORESTER

Original shape of train data frame: 5000 rows, 33 columns

NA values

There is no NA values in your data.

CREATING MODELS

--- Ranger model has been created ---
--- Catboost model has been created ---
--- Xgboost model has been created ---
--- LightGBM model has been created ---

COMPARISON

Results of compared models:

model	auc	recall	precision	f1	accuracy
-----	-----	-----	-----	-----	-----
Catboost	0.9995409	0.9990817	1.0000000	0.9995407	0.9993333
Ranger	0.5161192	1.0000000	0.7324702	0.8455790	0.7348333
XGboost	0.5000000	1.0000000	0.7260000	0.8412514	0.7260000
LightGBM	0.5000000	1.0000000	0.7260000	0.8412514	0.7260000

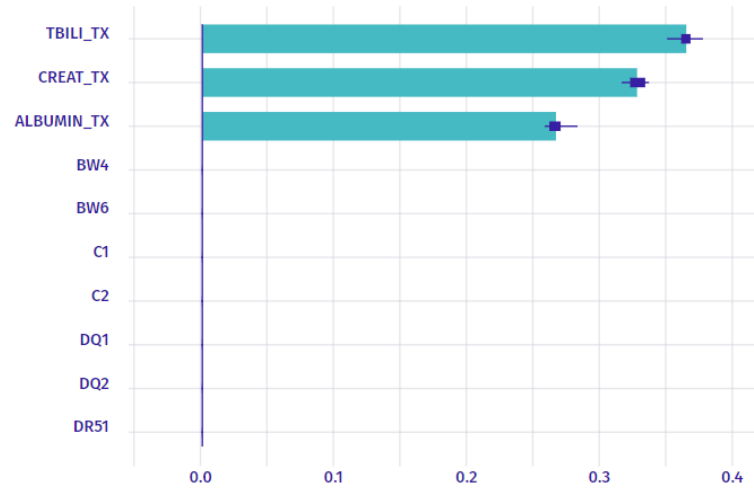
Interactive Studio for Catboost Model

ALBUMIN_TX

id: 2 | y: 0

Feature Importance

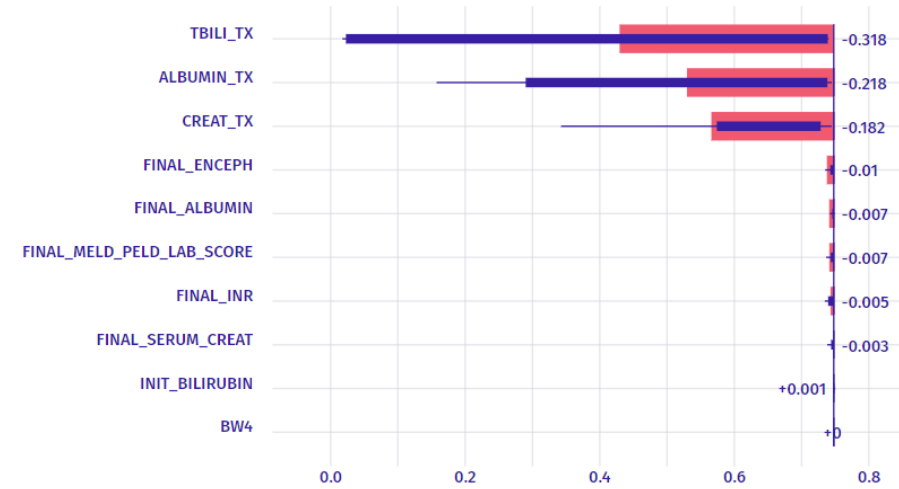
D X



One minus AUC

Shapley Values

D X



contribution

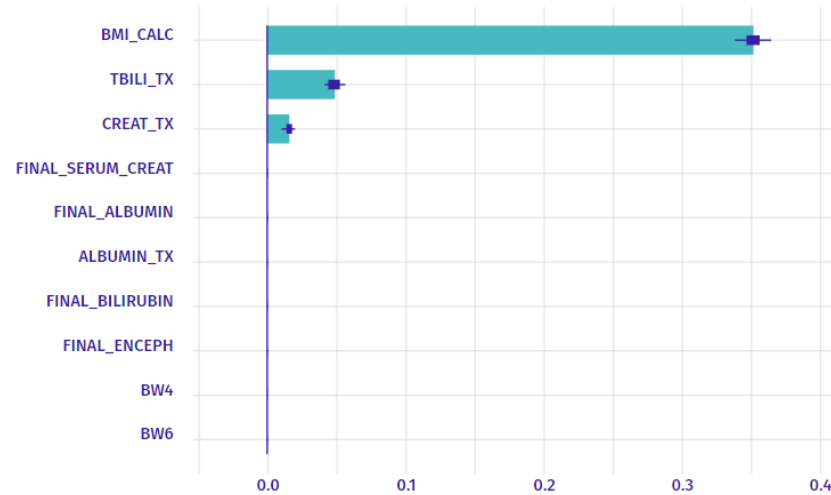
Interactive Studio for Catboost Model

ALBUMIN_TX

id: 1 | y: 0

Feature Importance

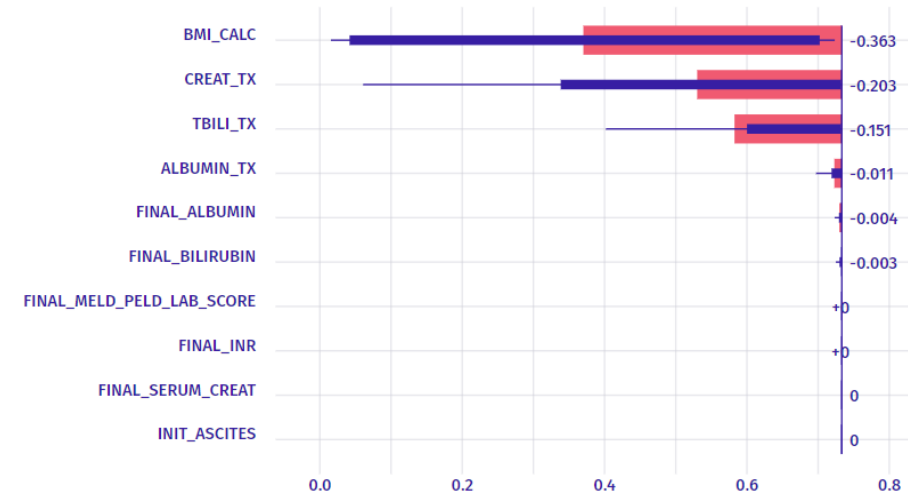
D X



One minus AUC

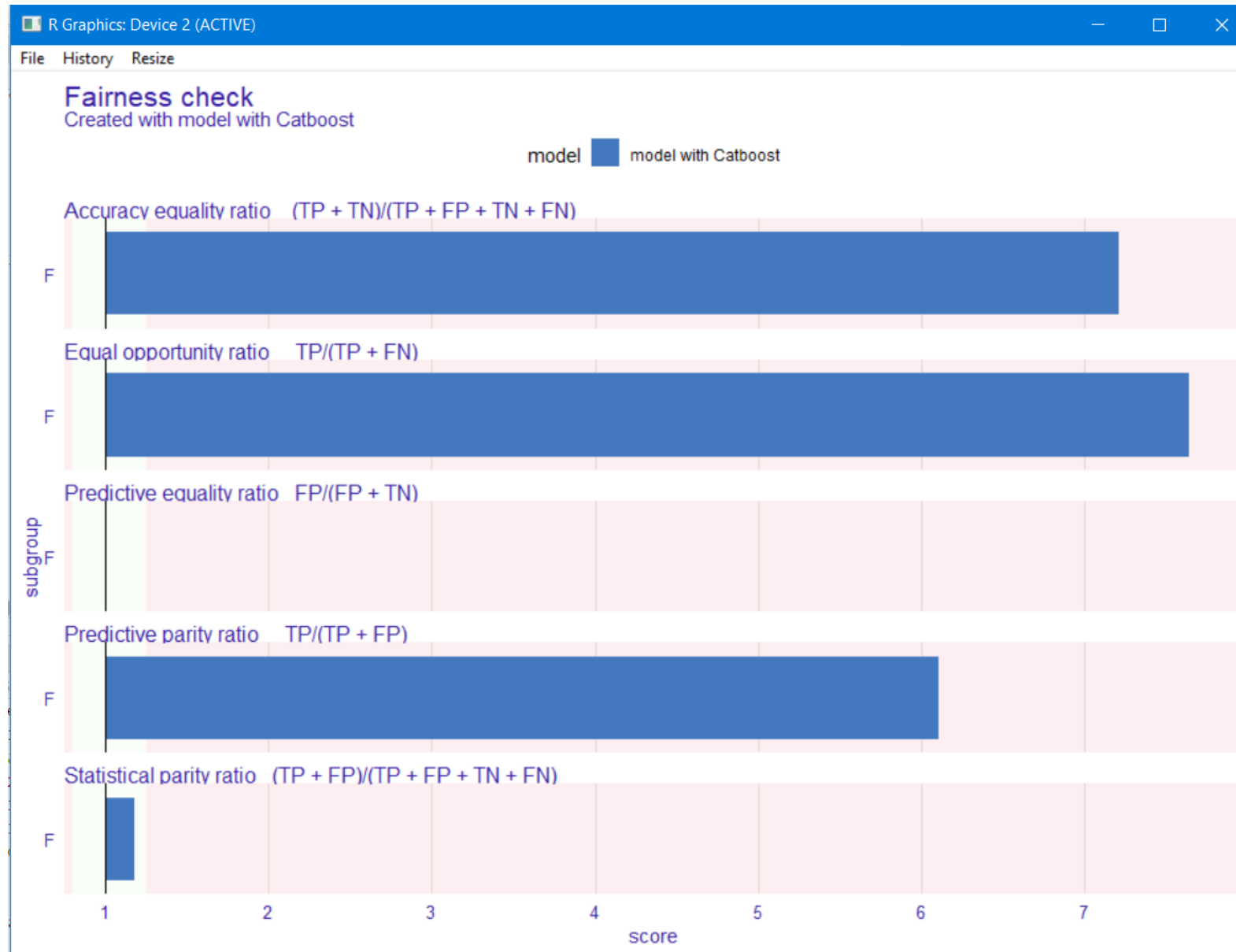
Shapley Values

D X



contribution

Fairness check



Conclusions

- Model returns almost perfect score for AUC (catboost)
- Removing demographic features (BMI-Calc, gender,...) and subjective features does not have a large effect on model performance
- Based on perturbation-important variables, we propose to add Albumin & BMI into MELD-Score model!
- Our model is unfair in respect to Gender, more biased to male.

References

- Predicting Mortality in Liver Transplant Candidates Explainable AI in Healthcare and Medicine, **2021**, Volume 914 ISBN : 978-3-030-53351-9 Jonathon Byrd, Sivaraman Balakrishnan, Xiaoqian Jiang, Zachary C. Lipton
- Mathematical
- Liver Transplant Waitlist, Part 1 | UCLA Transplantation Services
- <https://columbiasurgery.org/liver/liver-transplant-waiting-list>
- [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6902892/#:~:text=MELD%2DNa%20was%20calculated%20through,United%20Network%20for%20Organ%20Sharing\)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6902892/#:~:text=MELD%2DNa%20was%20calculated%20through,United%20Network%20for%20Organ%20Sharing)
- <https://www.organdonor.gov/learn/organ-donation-statistics>
- UNOS.org

Outline

1/ Transplant problem and sickest-first policy

2/ MELD-Score and optimistic results

3/ Applying ML Techniques

- Process of cleaning datasets
- Models and obtained results from a paper

4/ Related works and Q&A section

Discussion Time