

YAP470

TabNet: Attentive Interpretable Tabular Learning

Mustafa Mert Sandal & Murat Gencer

Terimler

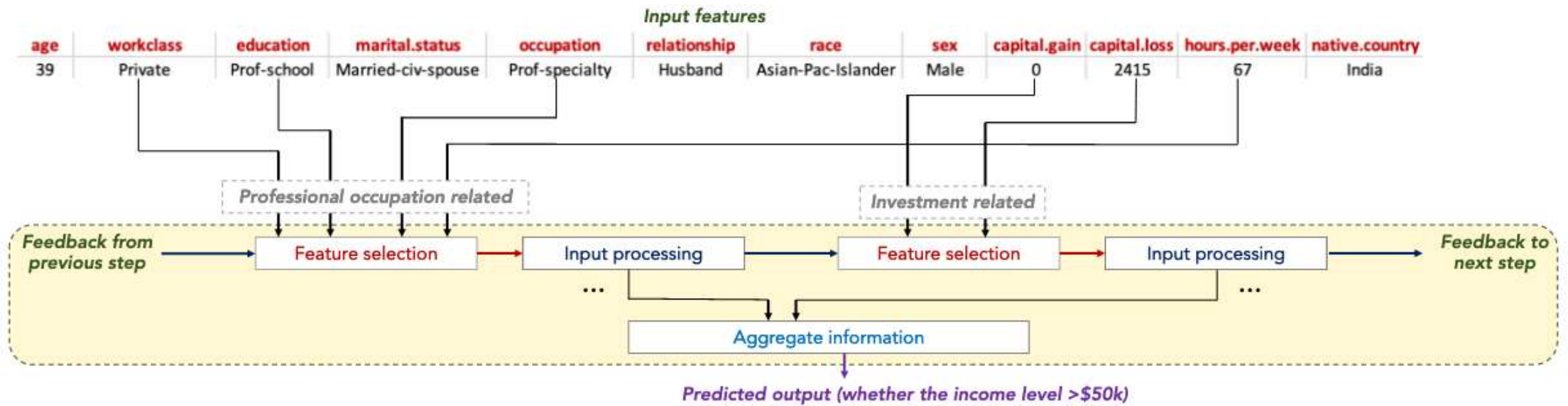
- Tabular data
- DNN(Deep Neural Network)
- Decision Tree

TABNET

- Tabnet tabular data için yeni bir DNN mimarisidir.

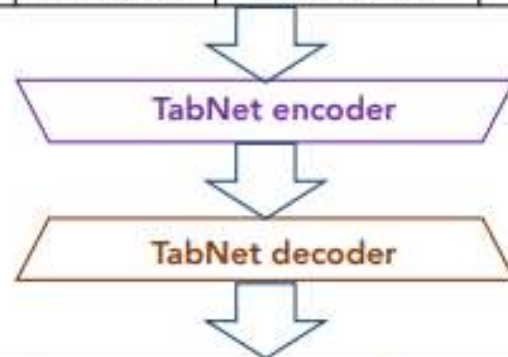
AVANTAJLARI

- Ön işleme adımları yok.
- Gradient Descent tabanlı eğitim
- Sequential Attention
- Yerel yorumlanabilirlik
- Genel yorumlanabilirlik.
- İyileştirme



Unsupervised pre-training

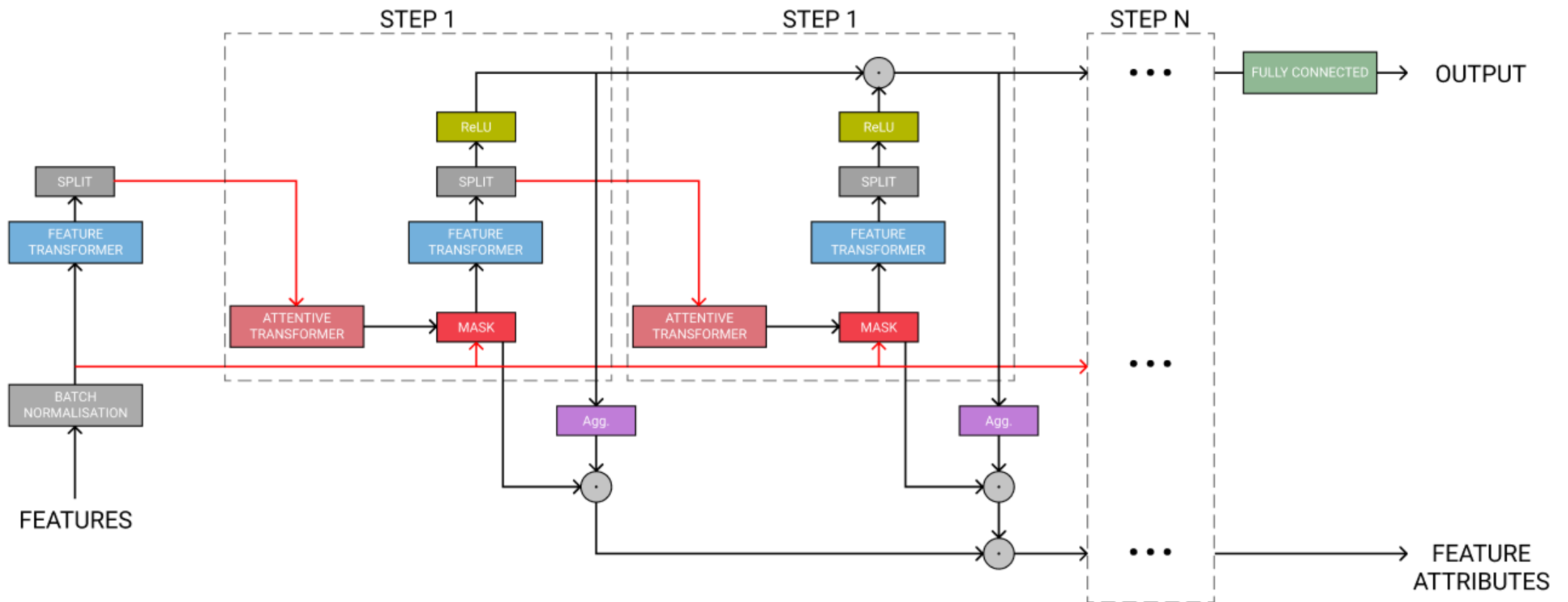
Age	Cap. gain	Education	Occupation	Gender	Relationship
53	200000	?	Exec-managerial	F	Wife
19	0	?	Farming-fishing	M	?
?	5000	Doctorate	Prof-specialty	M	Husband
25	?	?	Handlers-cleaners	F	Wife
59	300000	Bachelors	?	?	Husband
33	0	Bachelors	?	F	?
?	0	High-school	Armed-Forces	?	Husband

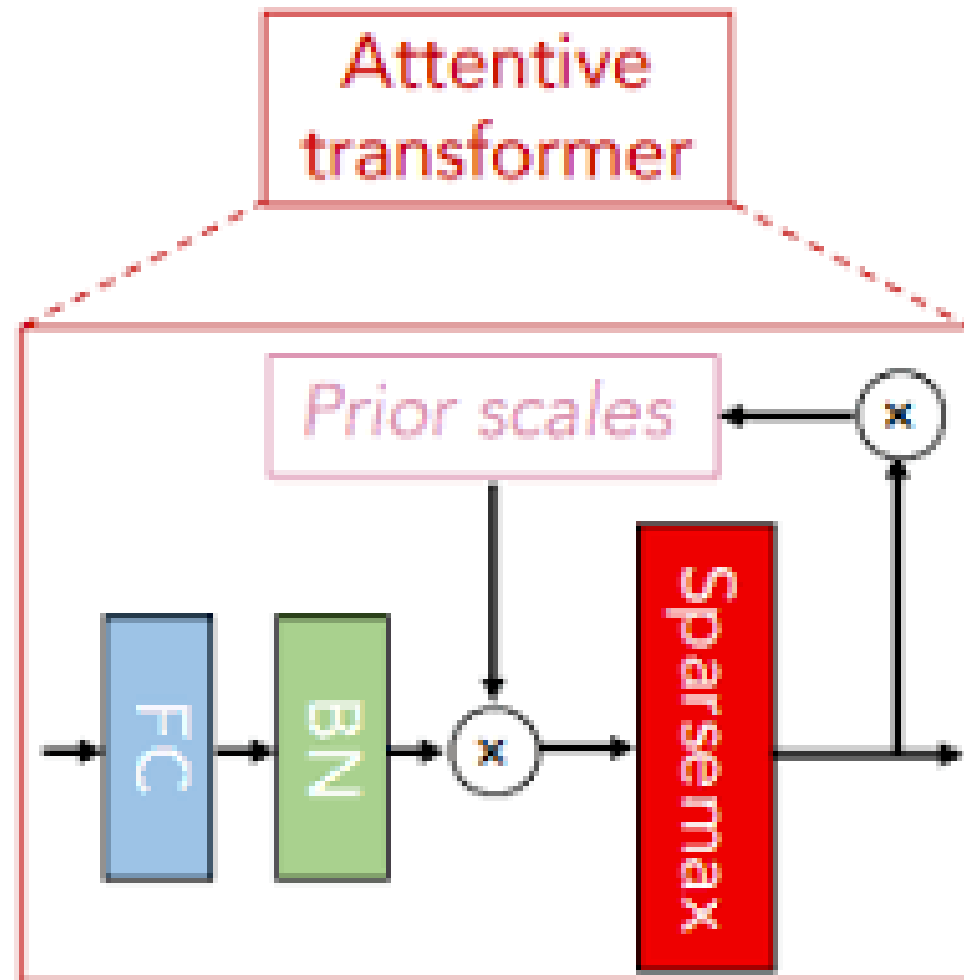


Age	Cap. gain	Education	Occupation	Gender	Relationship
		Masters			
		High-school			Unmarried
43					
	0	High-school		F	
			Exec-managerial	M	
			Adm-clerical		Wife
39				M	

- Feature Selection

TabNet Model Architecture





(d)

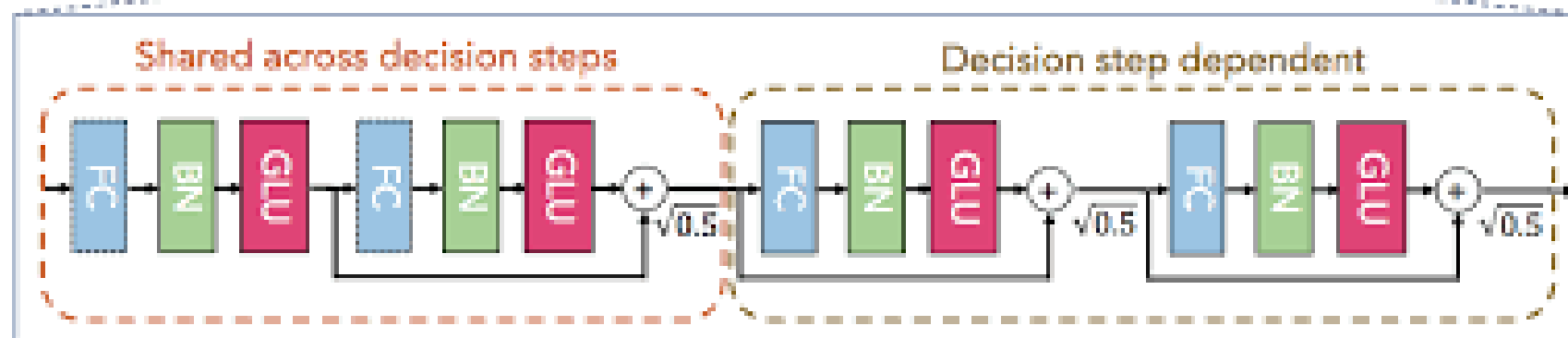
$$\mathbf{M}[\mathbf{i}] = \text{sparsemax}(\mathbf{P}[\mathbf{i} - \mathbf{1}] \cdot \mathbf{h}_i(\mathbf{a}[\mathbf{i} - \mathbf{1}]))$$

$$\mathbf{P}[\mathbf{i}] = \prod_{j=1}^i (\gamma - \mathbf{M}[\mathbf{j}])$$

Relaxation
Factor

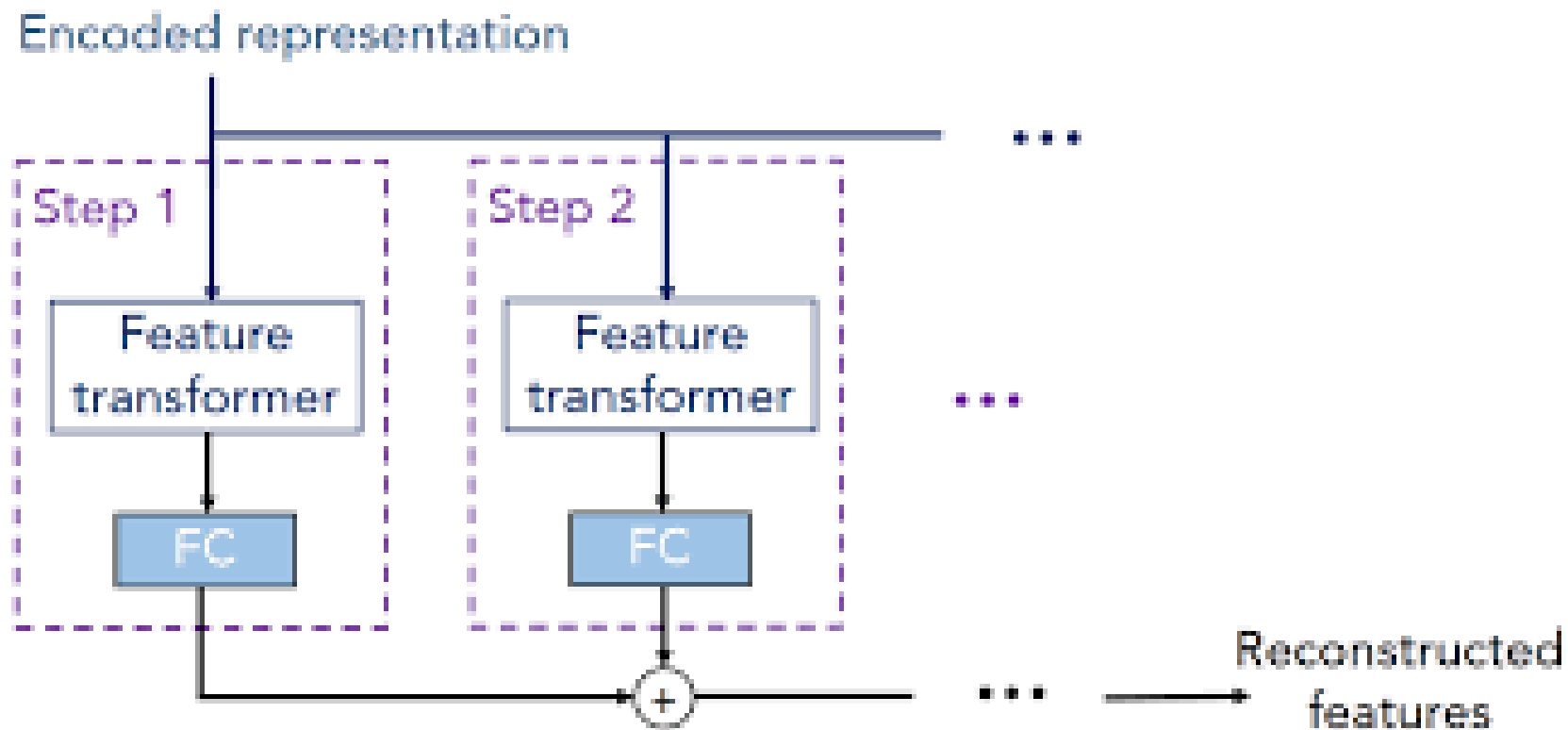
Previous Sparsemax
Activations

Feature
transformer



$$\mathbf{d}_{\text{out}} = \sum_{i=1}^{N_{steps}} \text{ReLU}(\mathbf{d}[\mathbf{i}])$$

$$[\mathbf{d}[\mathbf{i}], \mathbf{a}[\mathbf{i}]] = \mathbf{f}_i(\mathbf{M}[\mathbf{i}] \cdot \mathbf{f}), \text{ where } \mathbf{d}[\mathbf{i}] \in \mathbb{R}^{B \times N_d} \text{ and } \mathbf{a}[\mathbf{i}] \in \mathbb{R}^{B \times N_a}$$



$\mathbf{S} \in \{0, 1\}^{B \times D}$. The TabNet encoder inputs $(\mathbf{1} - \mathbf{S}) \cdot \hat{\mathbf{f}}$ and the decoder outputs the reconstructed features, $\mathbf{S} \cdot \hat{\mathbf{f}}$. We initialize $\mathbf{P}[0] = (\mathbf{1} - \mathbf{S})$ in encoder so that the model emphasizes merely on the known features, and the decoder's last FC layer is multiplied with \mathbf{S} to output the unknown features.

Deneyler

- TabNet, regression ve classification için inceleniyor.
- Sayısal sütunlar preprocessing olmadan kullanılıyor.
- Standard Classification (softmax cross entropy) ve Regression (MSE) loss fonksiyonları kullanılıyor.
- Alıntı yapılan tüm deneyler için, orijinal çalışma ile aynı train, validation ve test verileri kullanılıyor.

Instance-wise Feature Selection

- Özellikle küçük veri kümelerinde yüksek performans için çok önemlidir.
- 10k train örneğinden oluşan 6 tablo veri seti ele alınıyor.

<i>Model</i>	<i>Test AUC</i>					
	Syn1	Syn2	Syn3	Syn4	Syn5	Syn6
No selection	.578 ± .004	.789 ± .003	.854 ± .004	.558 ± .021	.662 ± .013	.692 ± .015
Tree	.574 ± .101	.872 ± .003	.899 ± .001	.684 ± .017	.741 ± .004	.771 ± .031
Lasso-regularized	.498 ± .006	.555 ± .061	.886 ± .003	.512 ± .031	.691 ± .024	.727 ± .025
L2X	.498 ± .005	.823 ± .029	.862 ± .009	.678 ± .024	.709 ± .008	.827 ± .017
INVASE	.690 ± .006	.877 ± .003	.902 ± .003	.787 ± .004	.784 ± .005	.877 ± .003
Global	.686 ± .005	.873 ± .003	.900 ± .003	.774 ± .006	.784 ± .005	.858 ± .004
<i>TabNet</i>	.682 ± .005	.892 ± .004	.897 ± .003	.776 ± .017	.789 ± .009	.878 ± .004

Gerçek Dünya Veri Kümelerinde Performans

- Forest Cover Type
- Poker Hand
- Sarcos
- Rossman Store Sales

Forest Cover Type (Orman Örtüsü Tipi)

- Amaç: Orman örtüsü türünün kartografik değişkenlerden sınıflandırılması.

<i>Model</i>	<i>Test accuracy (%)</i>
XGBoost	89.34
LightGBM	89.28
CatBoost	85.14
AutoML Tables	94.95
<i>TabNet</i>	96.99

Poker Hand (Poker Eli)

- Amaç: Kartların sıra ve renklerine göre poker elini sınıflandırmak.

<i>Model</i>	<i>Test accuracy (%)</i>
DT	50.0
MLP	50.0
Deep neural DT	65.1
XGBoost	71.1
LightGBM	70.0
CatBoost	66.6
<i>TabNet</i>	99.2
Rule-based	100.0

Sarcos

- Amaç: Antropomorfik bir robot kolunun ters dinamiklerini açıklamak.

<i>Model</i>	<i>Test MSE</i>	<i>Model size</i>
Random forest	2.39	16.7K
Stochastic DT	2.11	28K
MLP	2.13	0.14M
Adaptive neural tree	1.23	0.60M
Gradient boosted tree	1.44	0.99M
<i>TabNet-S</i>	1.25	6.3K
<i>TabNet-M</i>	0.28	0.59M
<i>TabNet-L</i>	0.14	1.75M

Rossman Store Sales (Rossman Mağaza Satışları)

- Amaç: Statik ve zamanla değişen özelliklerden mağaza satışlarını tahmin etmektir.

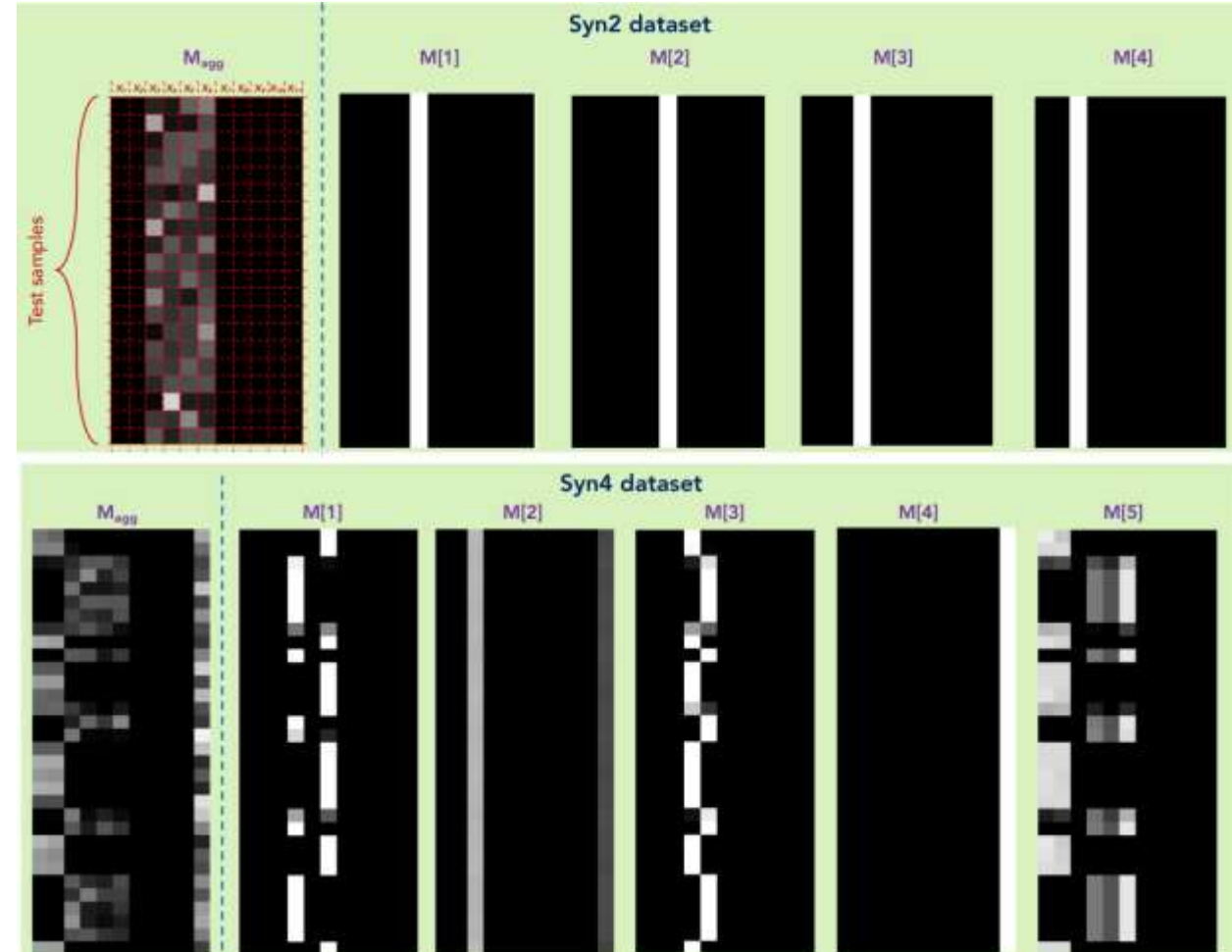
<i>Model</i>	<i>Test MSE</i>
MLP	512.62
XGBoost	490.83
LightGBM	504.76
CatBoost	489.75
<i>TabNet</i>	485.12

Yorumlanabilirlik

- Sentetik Veri Kümeleri
- Gerçek Dünya Veri Kümeleri

Sentetik Veri Kümeleri

Model	Test AUC					
	Syn1	Syn2	Syn3	Syn4	Syn5	Syn6
No selection	.578 ± .004	.789 ± .003	.854 ± .004	.558 ± .021	.662 ± .013	.692 ± .015
Tree	.574 ± .101	.872 ± .003	.899 ± .001	.684 ± .017	.741 ± .004	.771 ± .031
Lasso-regularized	.498 ± .006	.555 ± .061	.886 ± .003	.512 ± .031	.691 ± .024	.727 ± .025
L2X	.498 ± .005	.823 ± .029	.862 ± .009	.678 ± .024	.709 ± .008	.827 ± .017
INVASE	.690 ± .006	.877 ± .003	.902 ± .003	.787 ± .004	.784 ± .005	.877 ± .003
Global	.686 ± .005	.873 ± .003	.900 ± .003	.774 ± .006	.784 ± .005	.858 ± .004
TabNet	.682 ± .005	.892 ± .004	.897 ± .003	.776 ± .017	.789 ± .009	.878 ± .004



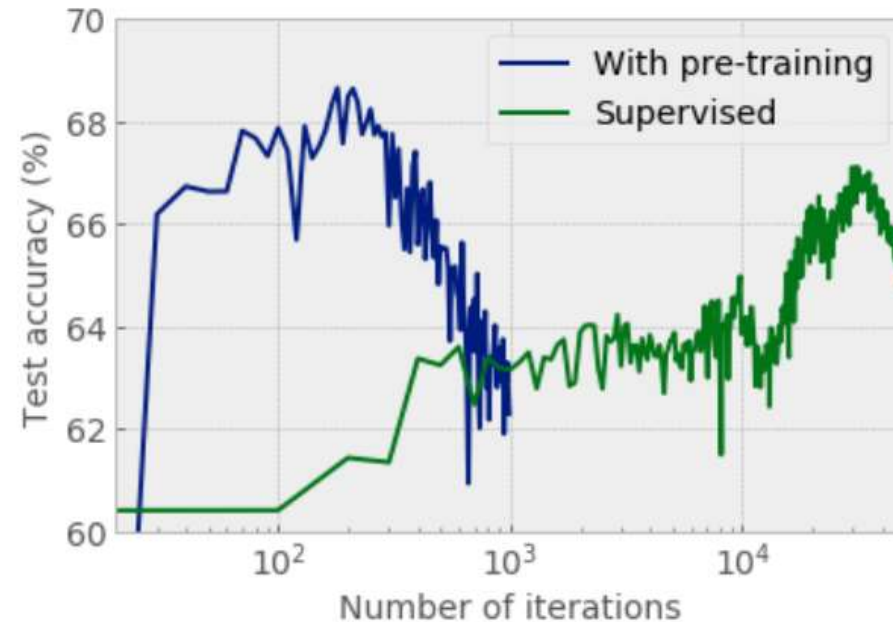
Gerçek Dünya Veri Kümeleri

- Mantar yenilebilirlik tahmini yapılmaktadır.
- Kokunun en ayırt edici özellik olduğu bilinmektedir.
- Yalnızca koku ile %98.5 test doğruluğu elde edilebilir.



Self-Supervised Learning (Kendi Kendini Denetleyen Öğrenme)

<i>Training dataset size</i>	<i>Test accuracy (%)</i>	
	<i>Supervised</i>	<i>With pre-training</i>
1k	57.47 ± 1.78	61.37 ± 0.88
10k	66.66 ± 0.88	68.06 ± 0.39
100k	72.92 ± 0.21	73.19 ± 0.15



Sonuç

- Tablolu öğrenme için yeni bir derin öğrenme mimarisi TabNet öğretiliyor.
- Instance-wise feature selection, en göze çarpan özellikler için kullanıldığından verimli öğrenme sağlıyor ve ayrıca selection mask'lerin görselleştirilmesi sayesinde daha yorumlanabilir bir hal alıyor.
- TabNet'in tabular veri kümelerinde önceki çalışmalardan daha iyi performans gösterdiği örnekleniyor.
- Son olarak, hızlı adaptasyon ve iyileştirilmiş performans için unsupervised pre-training'in önemli faydaları gösteriliyor.

Kaynakça

- **TabNet: Attentive Interpretable Tabular Learning**
- <https://towardsdatascience.com/tabnet-e1b979907694>
- <https://towardsdatascience.com/e2e-the-every-purpose-ml-method-5d4f20dafee4>
- <https://blogs.nvidia.com/blog/2020/05/14/sparsity-ai-inference/>