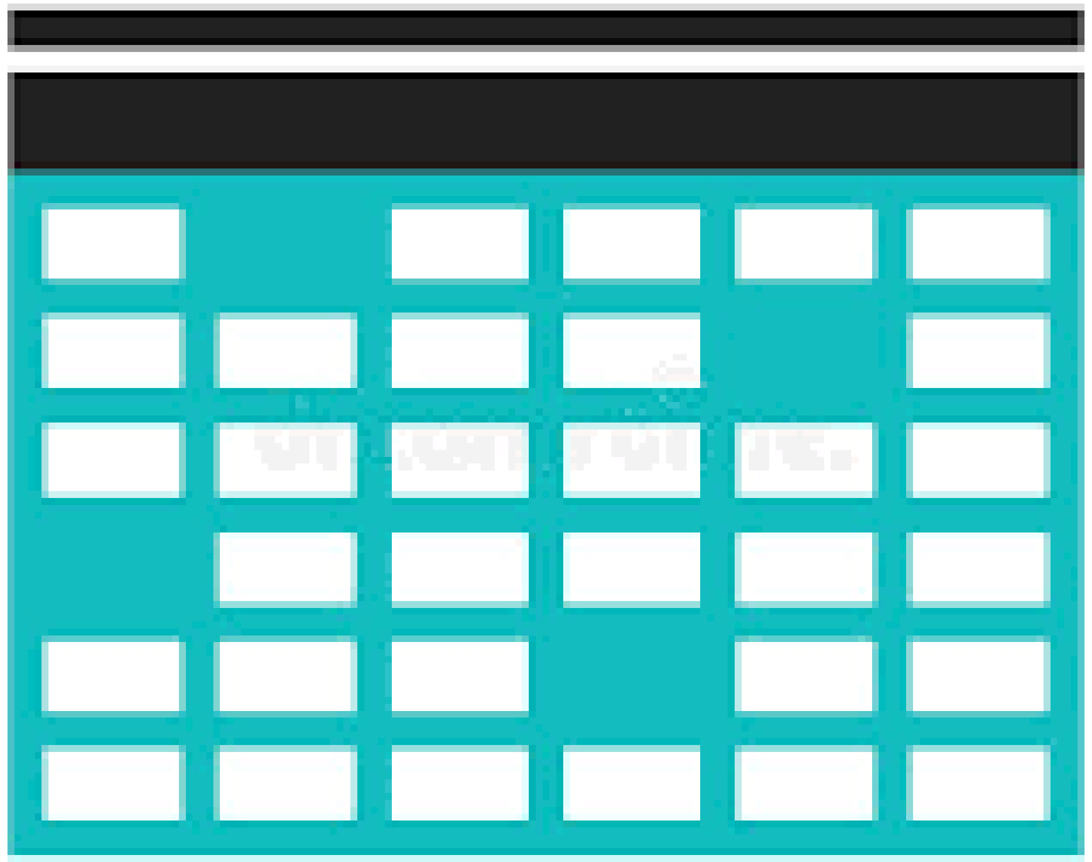


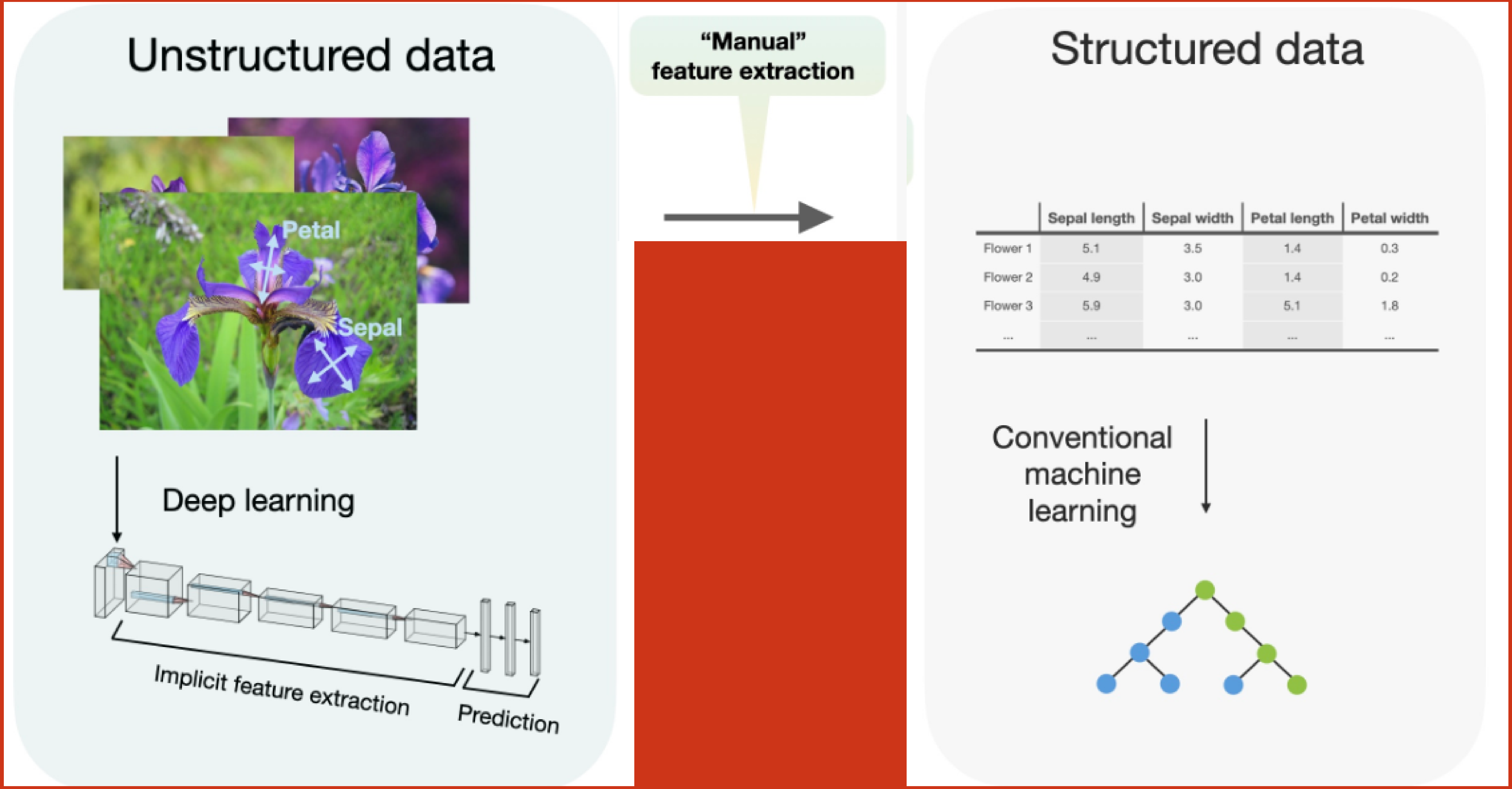
Overview of Deep Tabular Learning

TabNet/TabTransformer, Wide&Deep, VIME

What is Tabular Data?

Unstructured data

Structured data



	Sepal length	Sepal width	Petal length	Petal width
Flower 1	5.1	3.5	1.4	0.3
Flower 2	4.9	3.0	1.4	0.2
Flower 3	5.9	3.0	5.1	1.8
...	...	...	...	...

Problems of Tabular data for DL

1. Low-Quality Training Data
2. Missing or Complex Irregular Spatial Dependencies
3. Dependency on Preprocessing
4. Importance of Single Features

Deep learning approaches for tabular data

Transformer-based models

TAB NET

TabTransformer

Data encoding methods

VIME

SCARF

Regularization models

Hybrid models

wide & deep

TabNN

NON

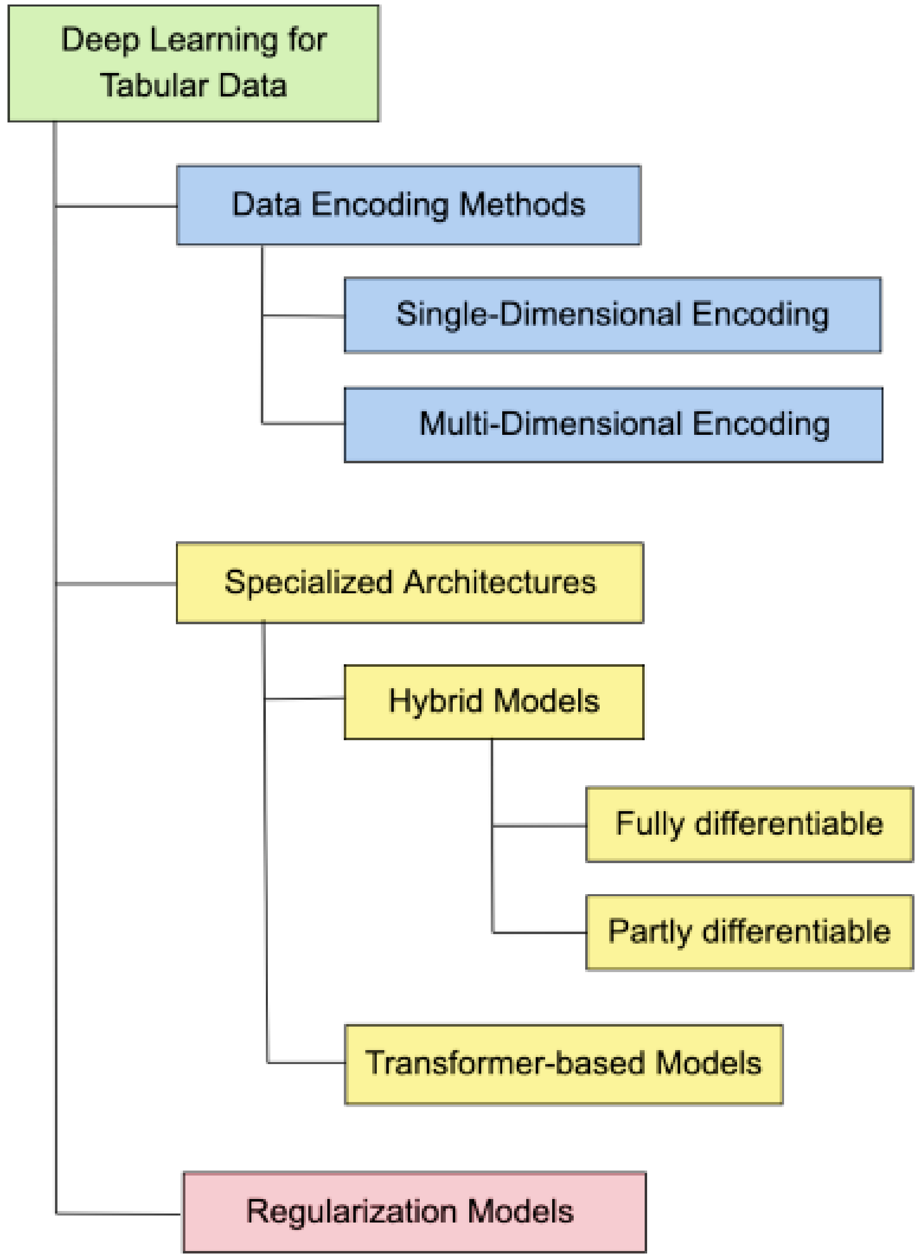


Figure 1: Unified taxonomy of deep neural network models for heterogeneous tabular data

Data encoding methods

Single-Dimensional Encoding

Height
Tall
Medium
Short

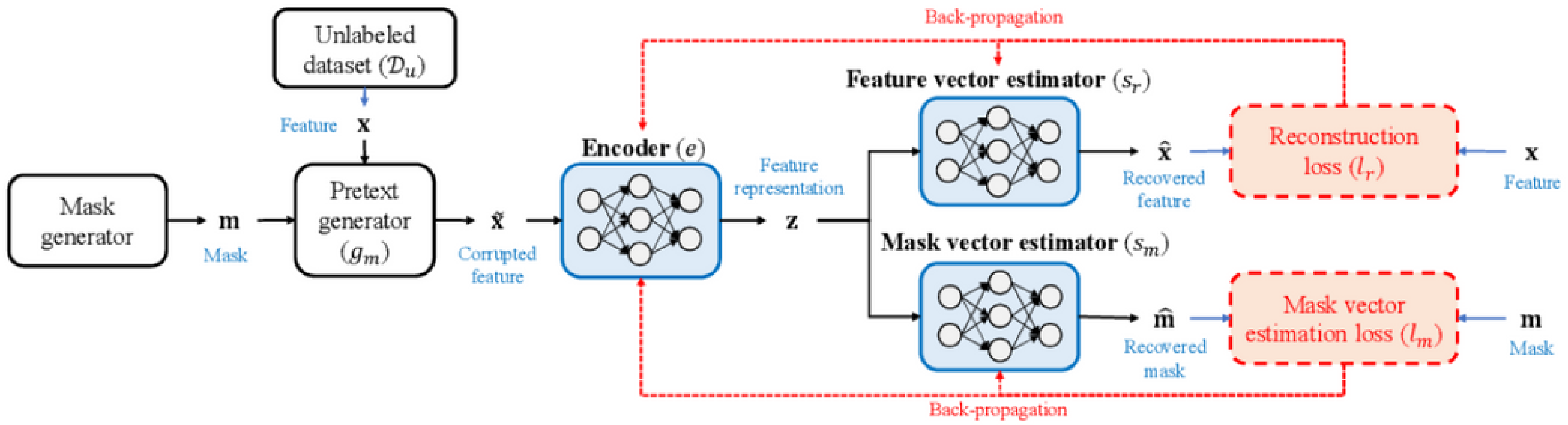
Height
0
1
2

Label encoding

Temperature	Order	Binary	Temperature_0	Temperature_1	Temperature_2
Hot	1	001	0	0	1
Cold	2	010	0	1	0
Very Hot	3	011	0	1	1
Warm	4	100	1	0	0
Hot	1	001	0	0	1
Warm	4	100	1	0	0
Warm	4	100	1	0	0
Hot	1	001	0	0	1
Hot	1	001	0	0	1
Cold	2	010	0	1	0

Binary encoding

Multi-Dimensional Encoding - > VIME

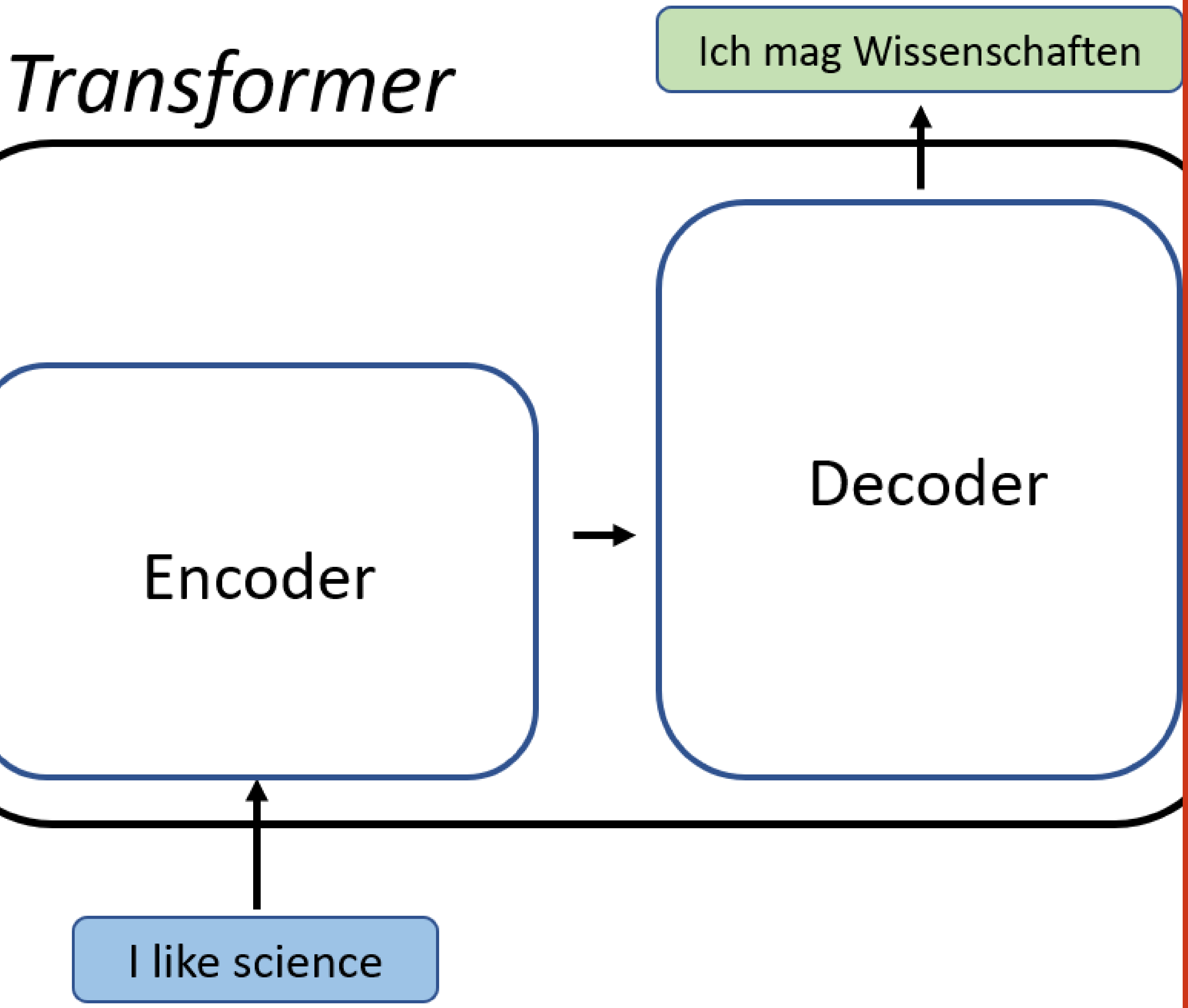


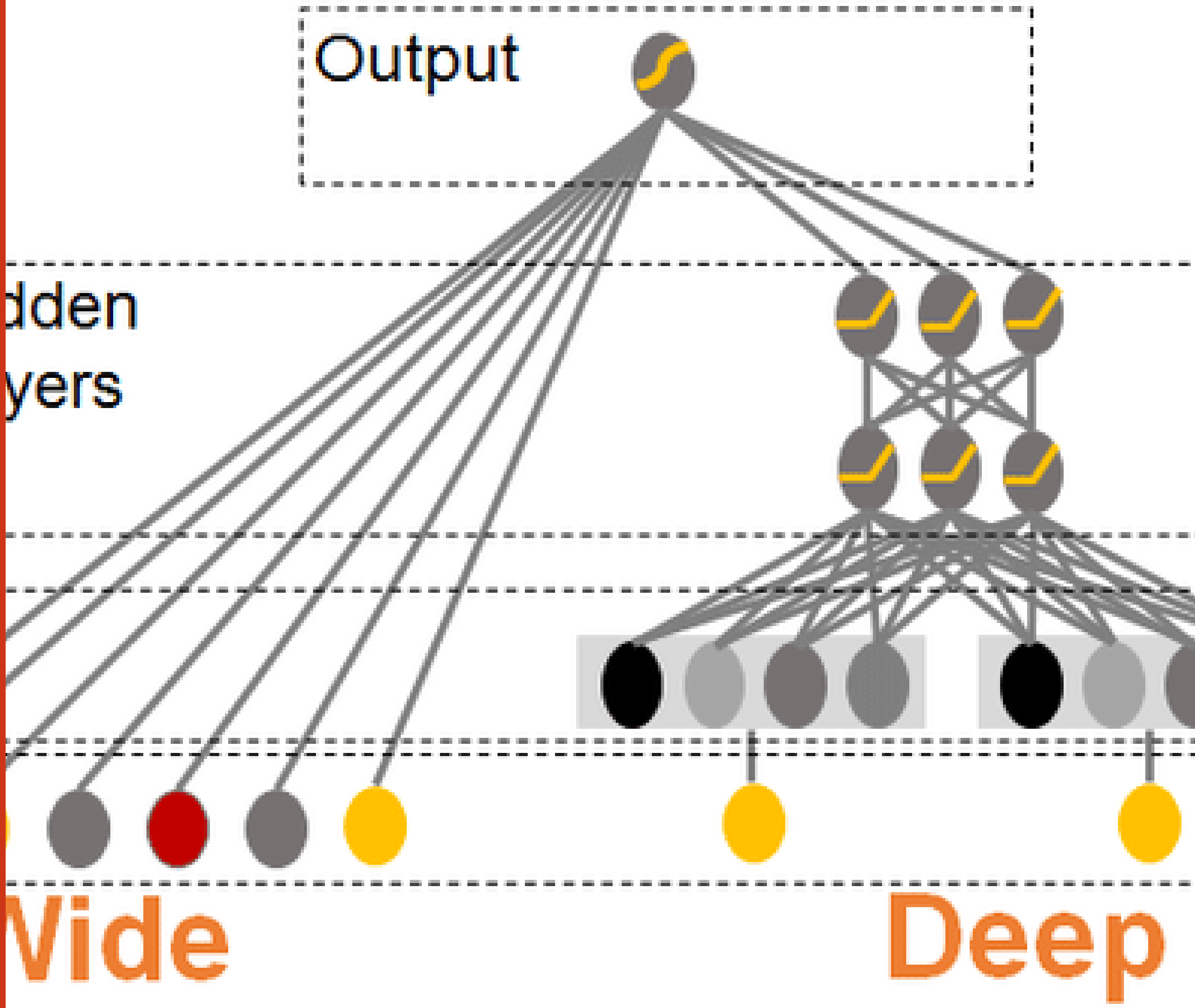
Total True (Survived = 1) of each class / Total Class

Pclass (X)	Survived (y)
Class 2	1
Class 3	0
Class 3	0
Class 1	1
Class 2	0

Transformed Pclass
0.5
0
0
1
0

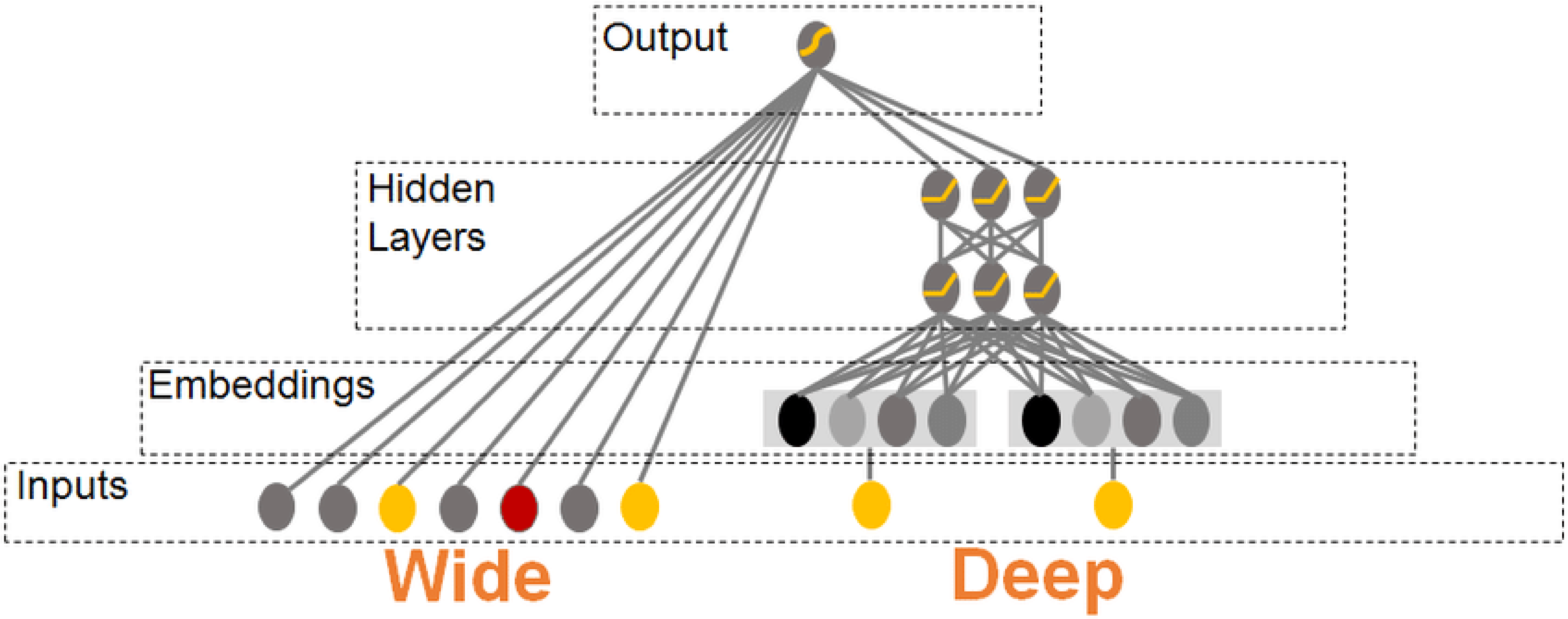
Leave One Out Encoding (LOOE)



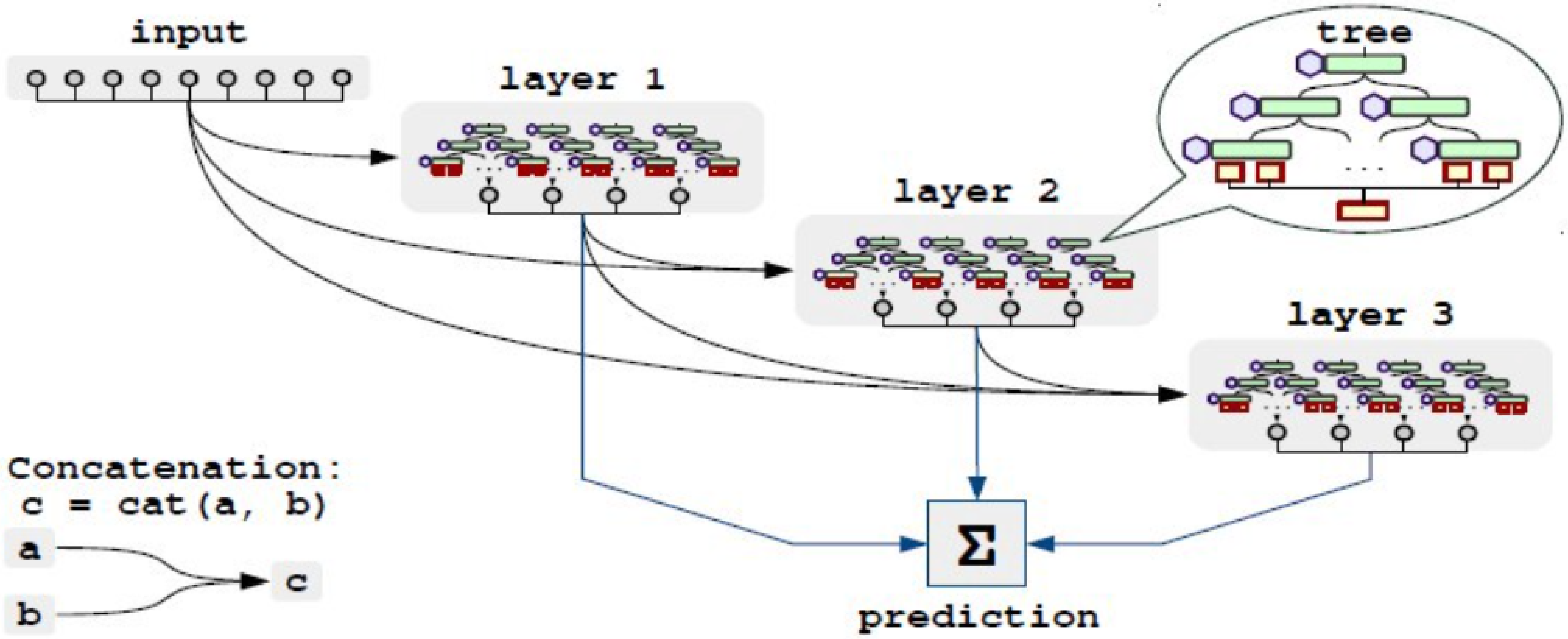


Specialized Architectures

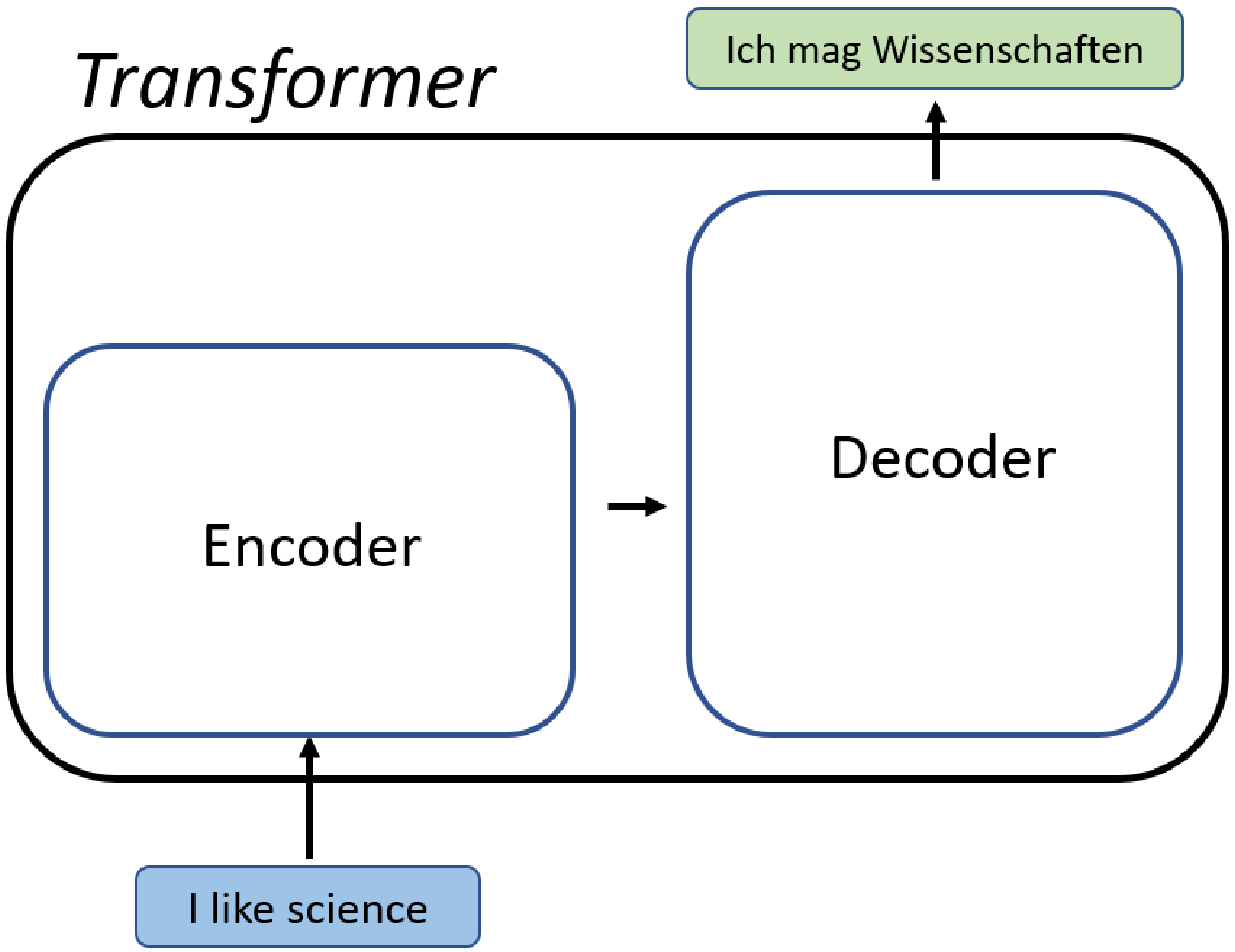
Specialized Architectures
Hybrid models - Wide&Deep



Specialized Architectures
Hybrid models - NODE

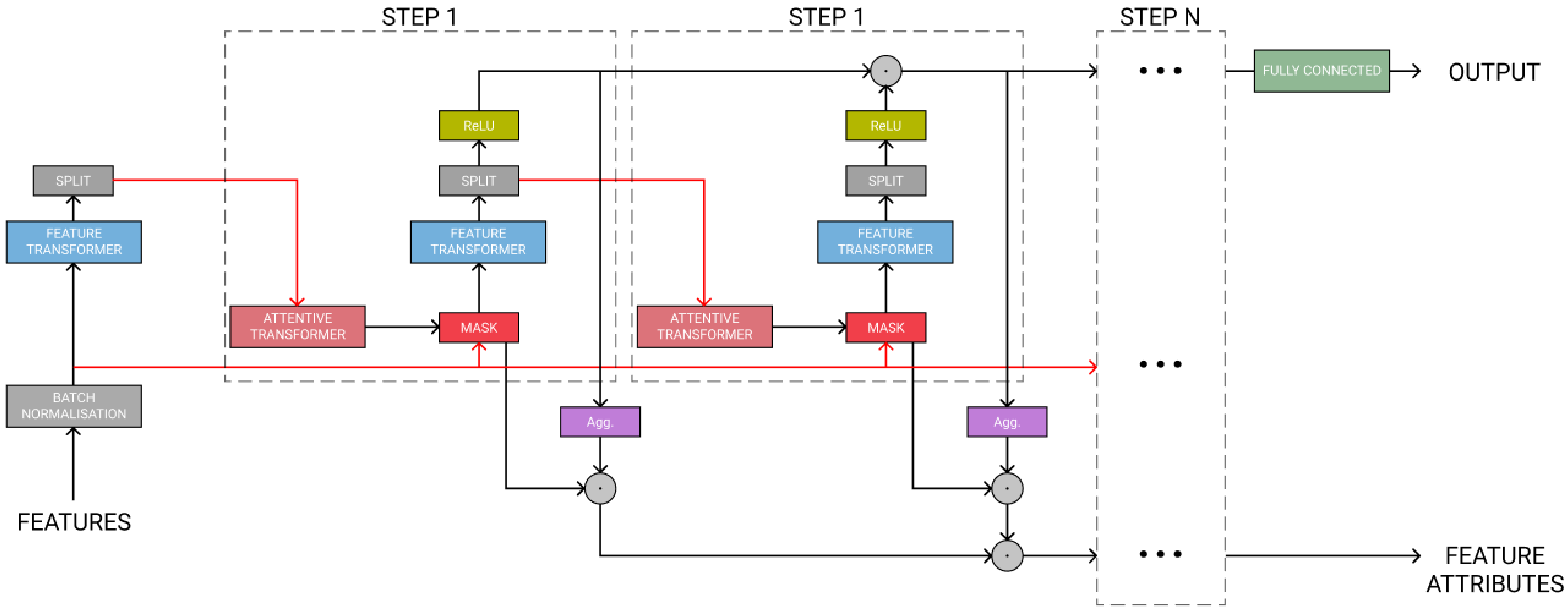


Specialized Architectures Transformer-based models

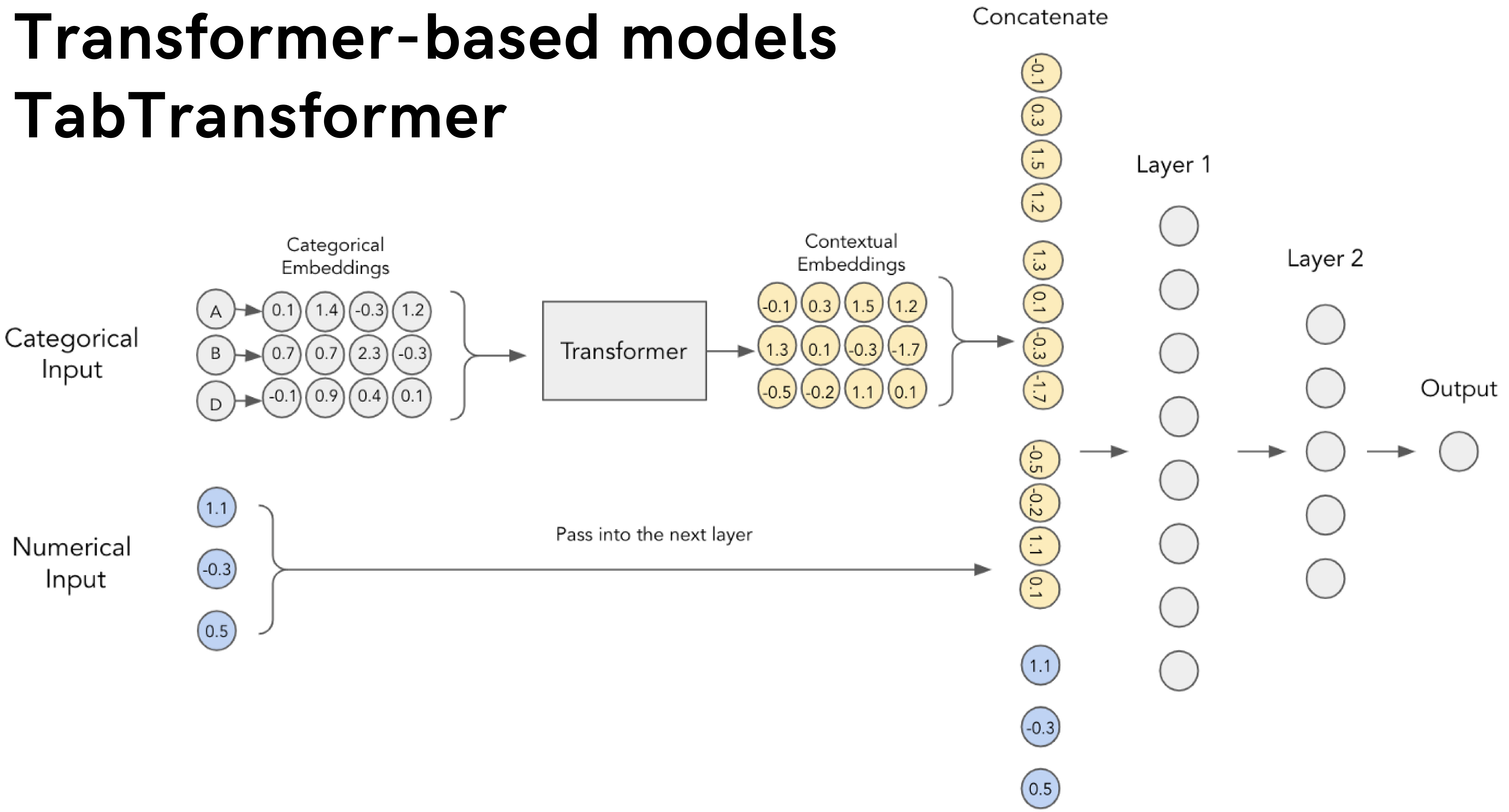


Transformer-based models -TabNet

TabNet Model Architecture



Transformer-based models
TabTransformer



Regularization models

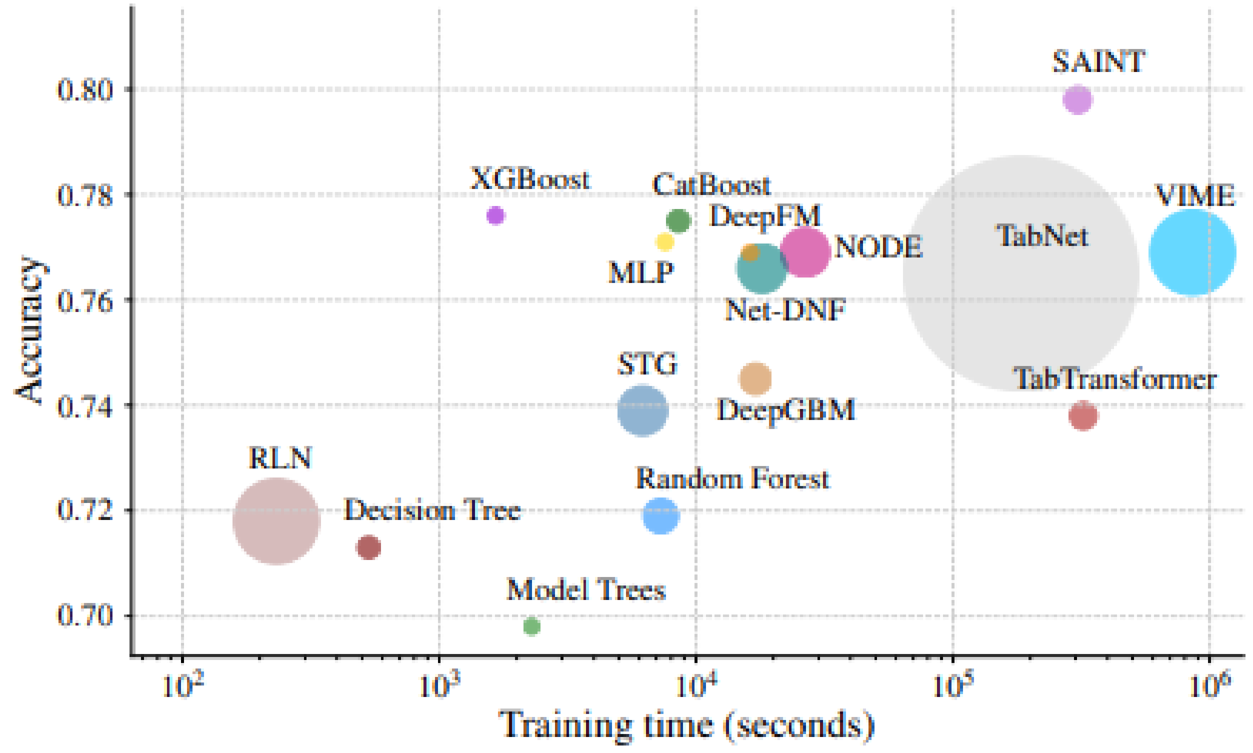
RLN - >"Counterfactual Loss".

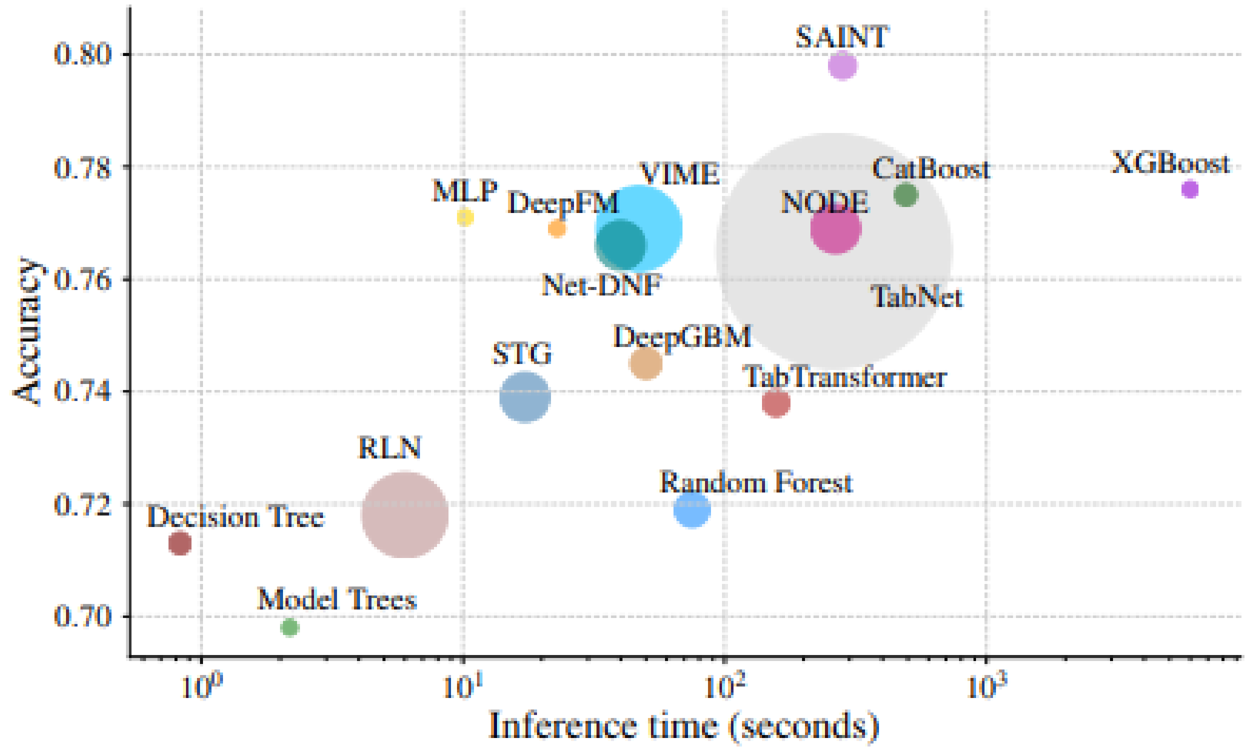
A "cocktail" of regularization - > thirteen different techniques

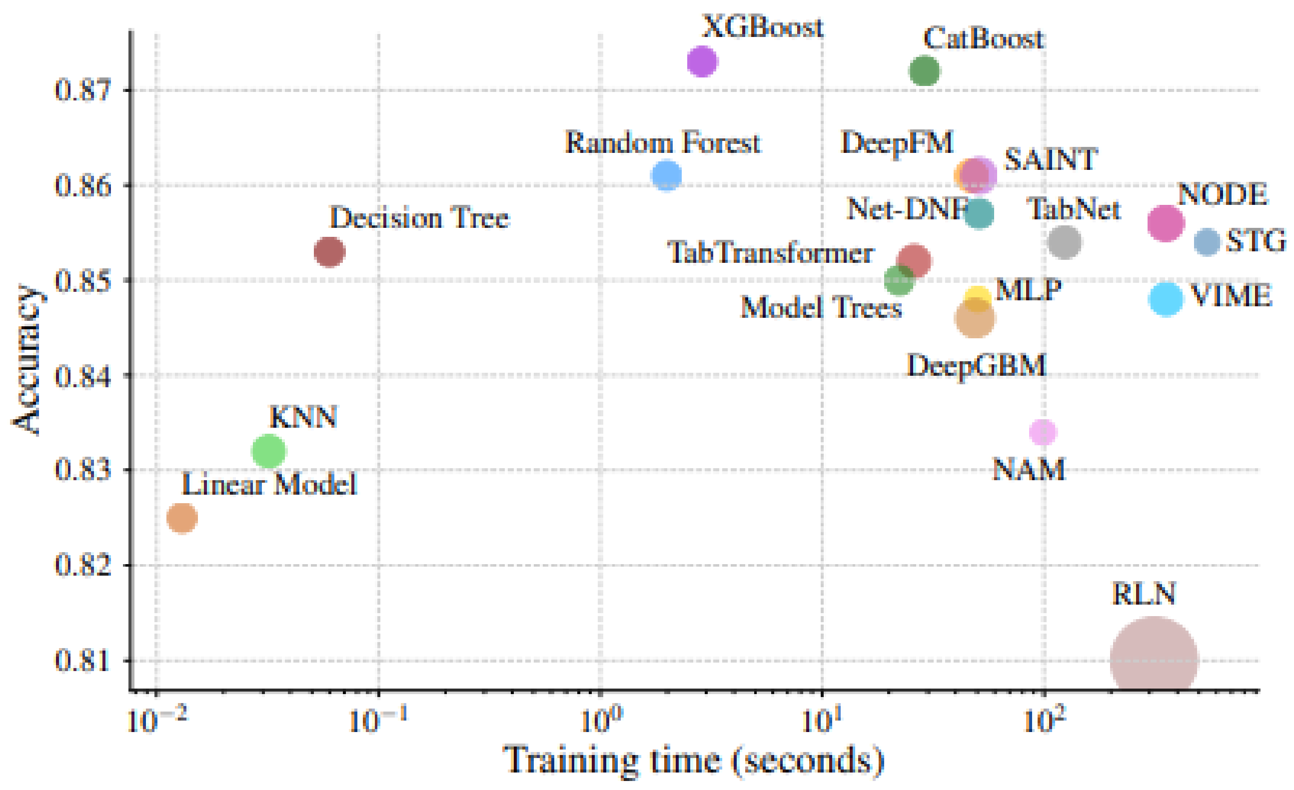
ML models vs DL models

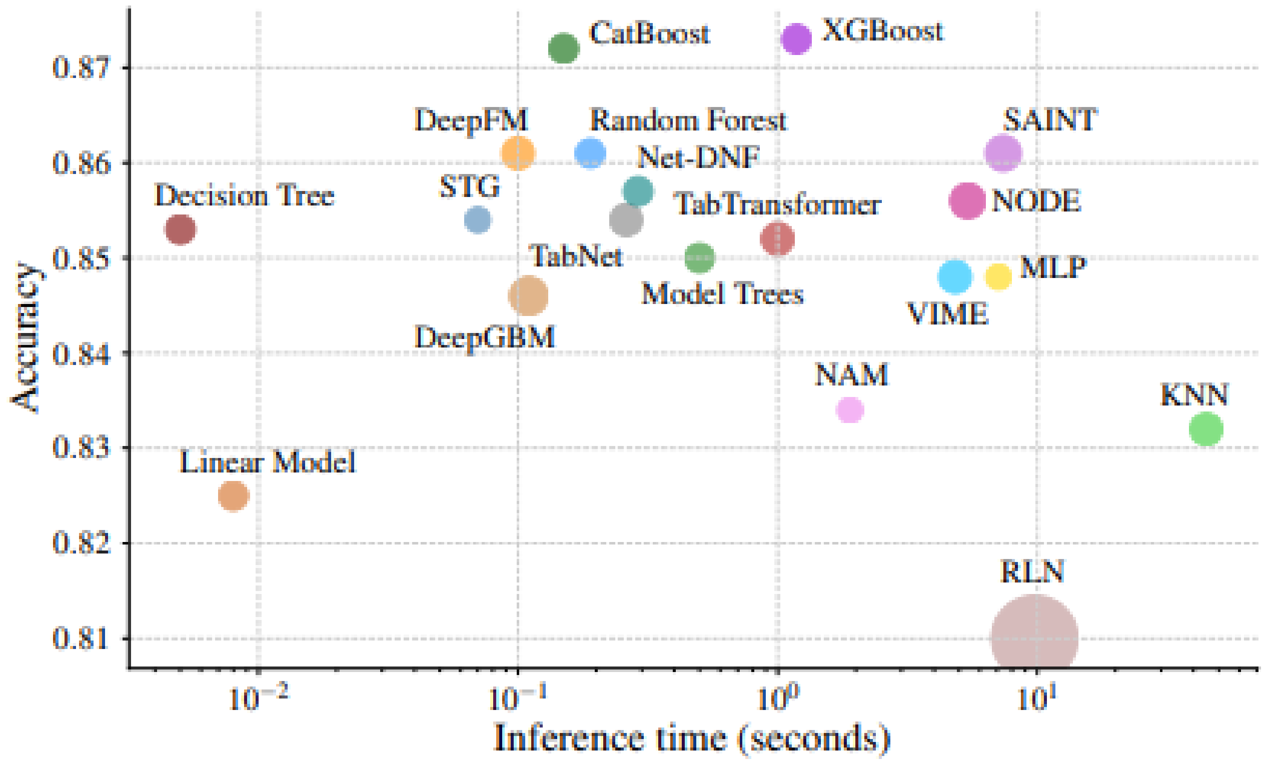
	HELOC		Adult		HIGGS		Covertype		Cal. Housing
	Acc ↑	AUC ↑	Acc ↑	AUC ↑	Acc ↑	AUC ↑	Acc ↑	AUC ↑	MSE ↓
Linear Model	73.0±0.0	80.1±0.1	82.5±0.2	85.4±0.2	64.1±0.0	68.4±0.0	72.4±0.0	92.8±0.0	0.528±0.008
KNN [65]	72.2±0.0	79.0±0.1	83.2±0.2	87.5±0.2	62.3±0.1	67.1±0.0	70.2±0.1	90.1±0.2	0.421±0.009
Decision Tree [197]	80.3±0.0	89.3±0.1	85.3±0.2	89.8±0.1	71.3±0.0	78.7±0.0	79.1±0.0	95.0±0.0	0.404±0.007
Random Forest [198]	82.1±0.2	90.0±0.2	86.1±0.2	91.7±0.2	71.9±0.0	79.7±0.0	78.1±0.1	96.1±0.0	0.272±0.006
XGBoost [53]	83.5±0.2	92.2±0.0	87.3±0.2	92.8±0.1	77.6±0.0	85.9±0.0	**97.3±0.0**	**99.9±0.0**	0.206±0.005
LightGBM [78]	83.5±0.1	92.3±0.0	**87.4±0.2**	**92.9±0.1**	77.1±0.0	85.5±0.0	93.5±0.0	99.7±0.0	**0.195±0.005**
CatBoost [79]	**83.6±0.3**	**92.4±0.1**	87.2±0.2	92.8±0.1	77.5±0.0	85.8±0.0	96.4±0.0	99.8±0.0	0.196±0.004
Model Trees [199]	82.6±0.2	91.5±0.0	85.0±0.2	90.4±0.1	69.8±0.0	76.7±0.0	-	-	0.385±0.019
MLP [200]	73.2±0.3	80.3±0.1	84.8±0.1	90.3±0.2	77.1±0.0	85.6±0.0	91.0±0.4	76.1±3.0	0.263±0.008
DeepFM [15]	73.6±0.2	80.4±0.1	86.1±0.2	91.7±0.1	76.9±0.0	83.4±0.0	-	-	0.260±0.006
DeepGBM [70]	78.0±0.4	84.1±0.1	84.6±0.3	90.8±0.1	74.5±0.0	83.0±0.0	-	-	0.856±0.065
RLN [72]	73.2±0.4	80.1±0.4	81.0±1.6	75.9±8.2	71.8±0.2	79.4±0.2	77.2±1.5	92.0±0.9	0.348±0.013
TabNet [5]	81.0±0.1	90.0±0.1	85.4±0.2	91.1±0.1	76.5±1.3	84.9±1.4	93.1±0.2	99.4±0.0	0.346±0.007
VIME [88]	72.7±0.0	79.2±0.0	84.8±0.2	90.5±0.2	76.9±0.2	85.5±0.1	90.9±0.1	82.9±0.7	0.275±0.007
TabTransformer [98]	73.3±0.1	80.1±0.2	85.2±0.2	90.6±0.2	73.8±0.0	81.9±0.0	76.5±0.3	72.9±2.3	0.451±0.014
NODE [6]	79.8±0.2	87.5±0.2	85.6±0.3	91.1±0.2	76.9±0.1	85.4±0.1	89.9±0.1	98.7±0.0	0.276±0.005
Net-DNF [57]	82.6±0.4	91.5±0.2	85.7±0.2	91.3±0.1	76.6±0.1	85.1±0.1	94.2±0.1	99.1±0.0	-
STG [201]	73.1±0.1	80.0±0.1	85.4±0.1	90.9±0.1	73.9±0.1	81.9±0.1	81.8±0.3	96.2±0.0	0.285±0.006
NAM [202]	73.3±0.1	80.7±0.3	83.4±0.1	86.6±0.1	53.9±0.6	55.0±1.2	-	-	0.725±0.022
SAINT [9]	82.1±0.3	90.7±0.2	86.1±0.3	91.6±0.2	**79.8±0.0**	**88.3±0.0**	96.3±0.1	99.8±0.0	0.226±0.004

ML models vs DL models









Summary

1. The best solutions for tabular data is still Decision Tree Ensembles (XGBoost, LightGBM, and Cat Boost)

2. For very large data sets, approaches based on deep learning may still be able to achieve competitive performance and even outperform classical models.

3. There is a necessity for a standardized benchmarking procedure

Summary

1. The best solutions for tabular data is still Decision Tree Ensembles (XGBoost, LightGBM, and Cat Boost)

2. For very large data sets, approaches based on deep learning may still be able to achieve competitive performance and even outperform classical models.

3. There is a necessity for a standardized benchmarking procedure

Articles:

Deep Neural Networks and Tabular Data: A Survey

https://arxiv.org/pdf/2110.01889.pdf

Revisiting Deep Learning Models for Tabular Data

https://arxiv.org/pdf/2106.11959.pdf

A Short Chronology Of Deep Learning For Tabular Data

https://analyticsindiamag.com/creating-deep-learning-models-for-tabular-data-using-rtdl/

Thanks for your attention!

– Tazhibay Aizhan