Facilitating Machine Learning Tasks Through Visual Analytics:

A Multi-perspective View

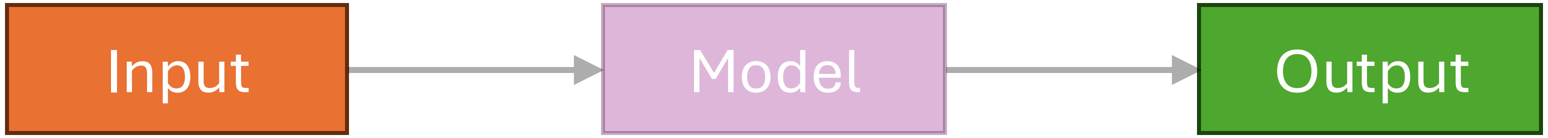
Linhao Meng
l.meng1@tue.nl

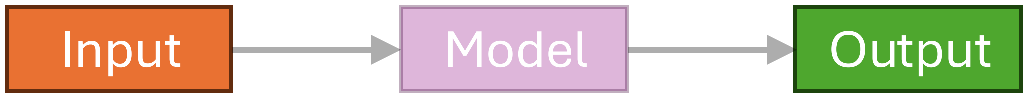
Eindhoven University of Technology
Visualization cluster

ELLIIT
Excellence Center at Linköping – Lund in Information Technology

Main Perspectives in ML Workflow

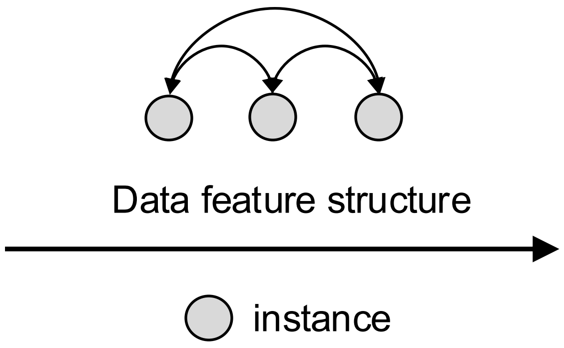
Main Perspectives in ML Workflow



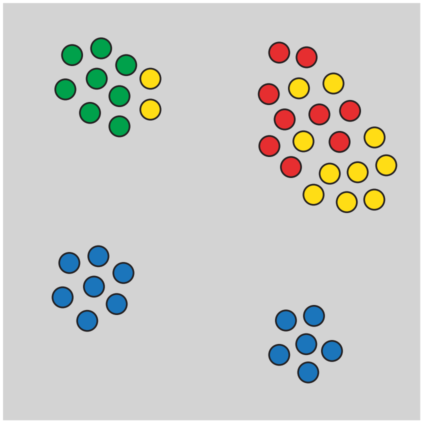


Data features

	Feature1	Feature2	Feature3	Feature4
1	25	M	0	True
2	50	F	1	False
3	30	M	2	True
4	20	F	0	True
…	…	…	…	…

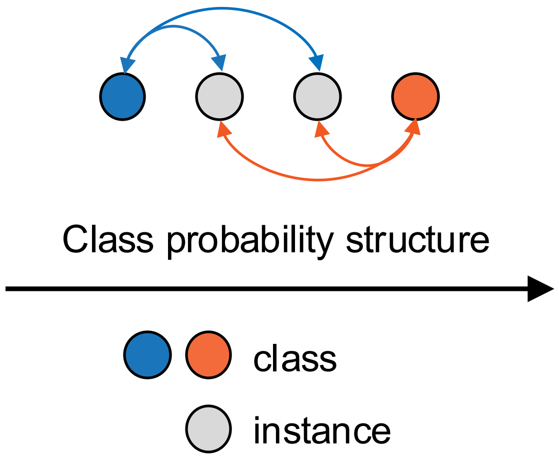


Feature Projection

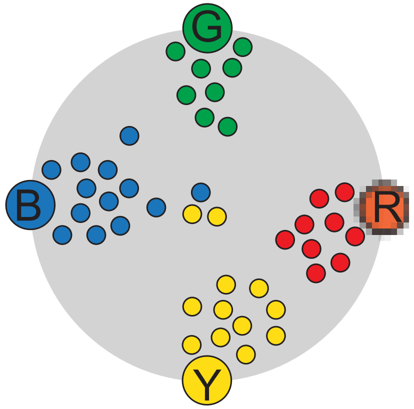


Class Probabilities

	Class-G	Class-B	Class-Y	Class-R
1	0.45	0	0.55	0
2	0.7	0.2	0	0.1
3	0	0.5	0	0.5
4	0.22	0.25	0.28	0.25
…	…	…	…	…



Class Radial Vis

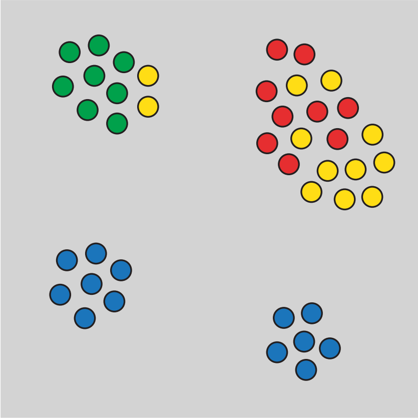


(Seifert et al. IV 2009)

Input → Model → Output

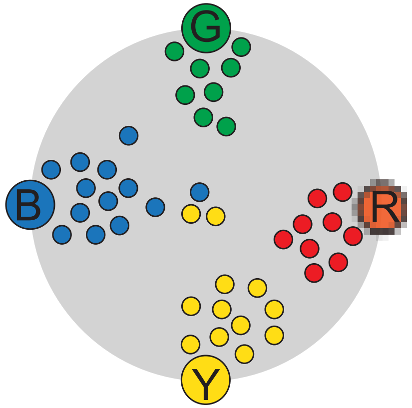
Standard Approach for Comparison

Feature Projection



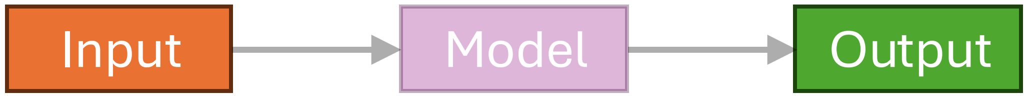
feature

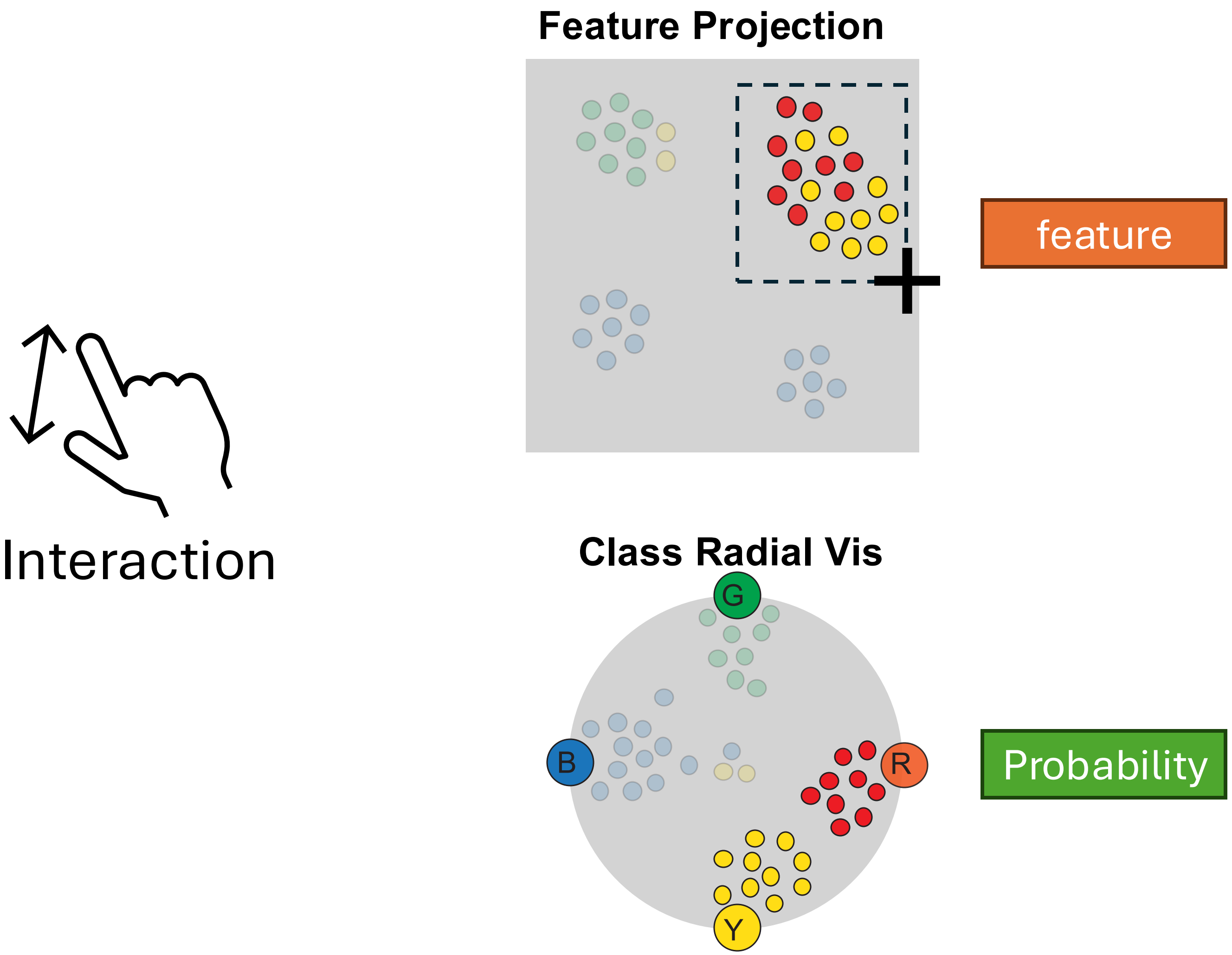
Class Radial Vis



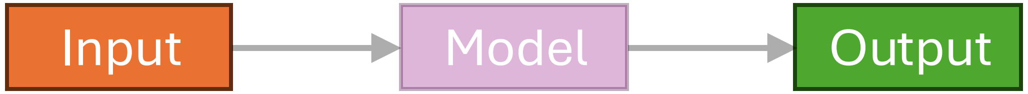
Probability

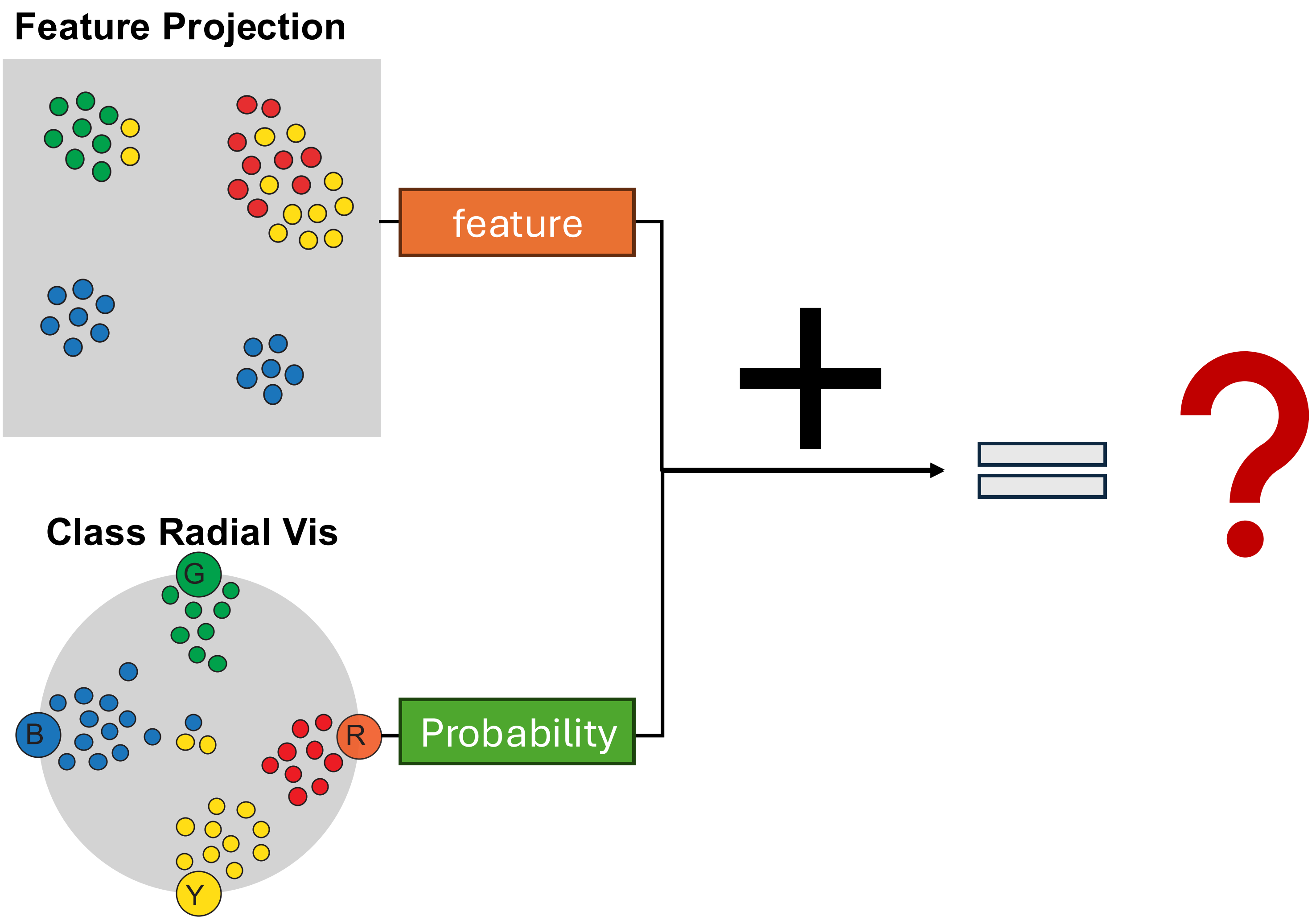
Standard Approach for Comparison





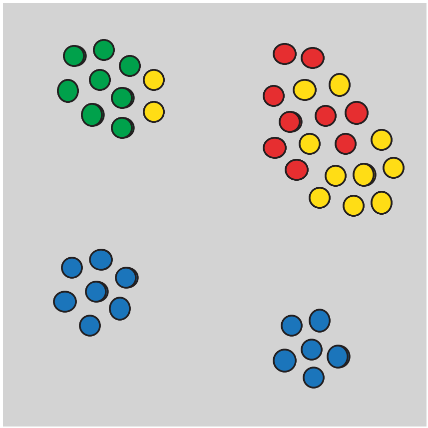
An alternative approach

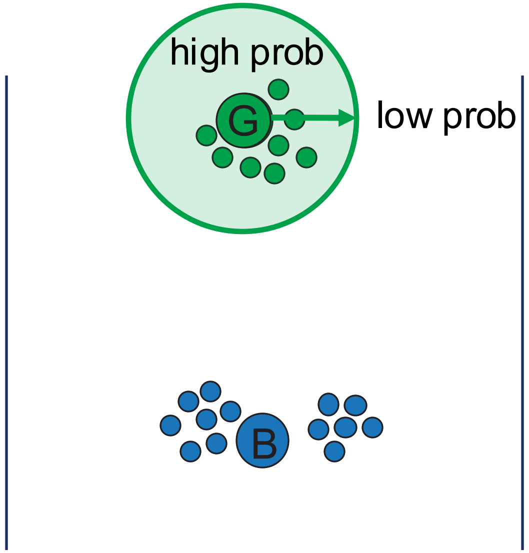




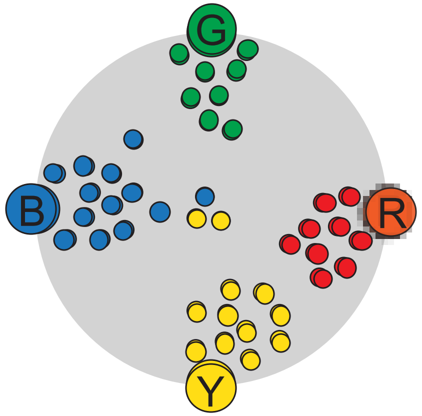
Combination

Feature Projection





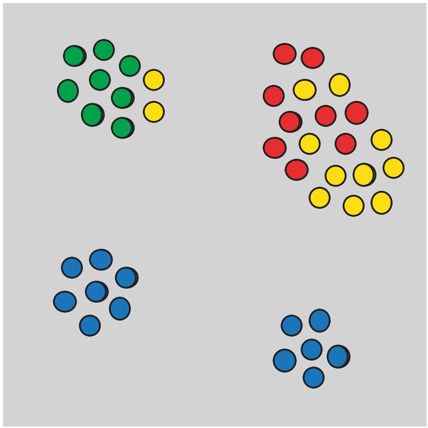
Class Radial Vis

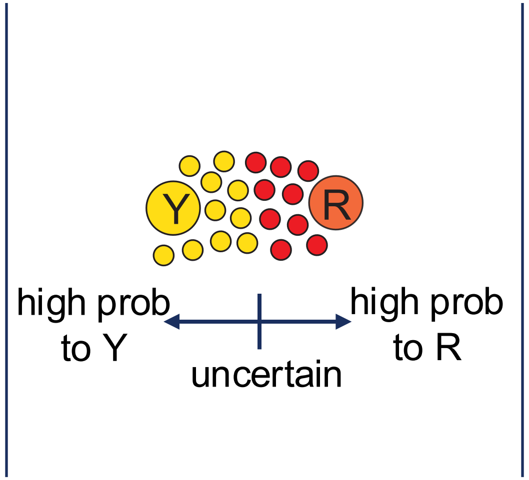


- ✓ Preservation of significant data feature structure
- ✓ Intra-cluster arrangement for revealing class probabilities

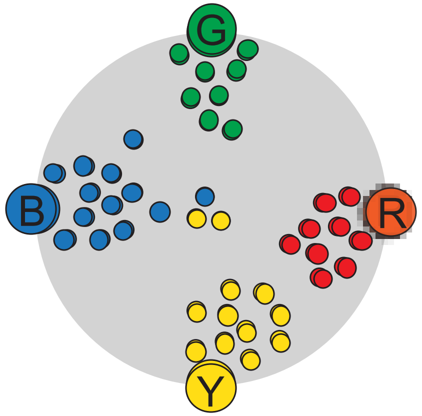
Combination

Feature Projection



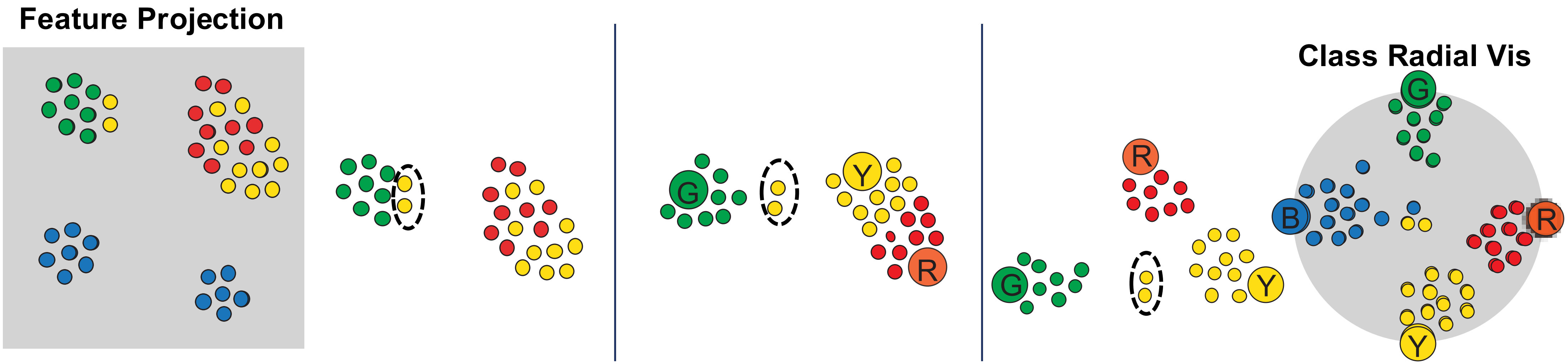


Class Radial Vis



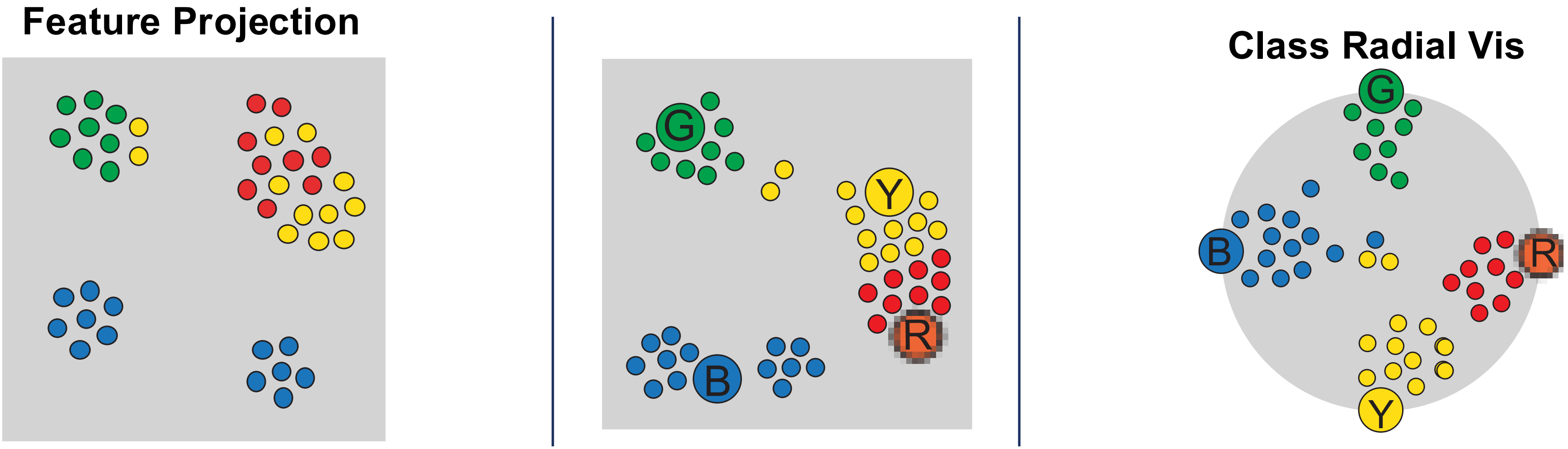
✓ Improved separation in ambiguous clusters

Combination



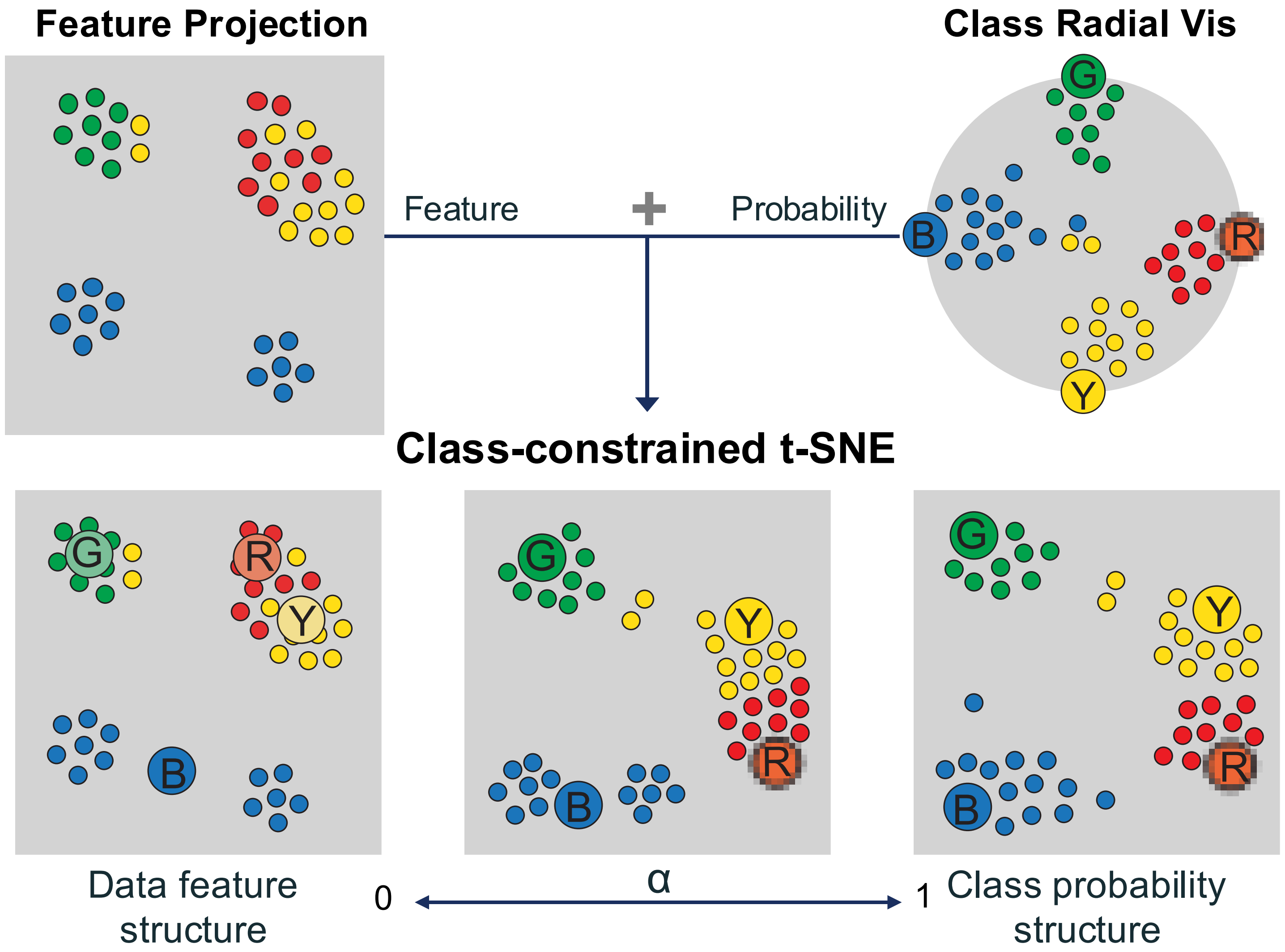
✓ Outliers pushed out from the main clusters

Class-constrained t-SNE

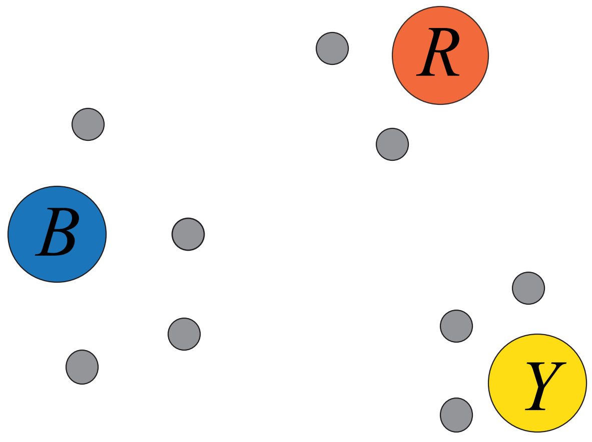


Class-constrained t-SNE

https://github.com/alicelh/class-constrained-t-SNE



Method



Data feature structure

Cost function: $C_d = C1 + C2$

Class probability structure

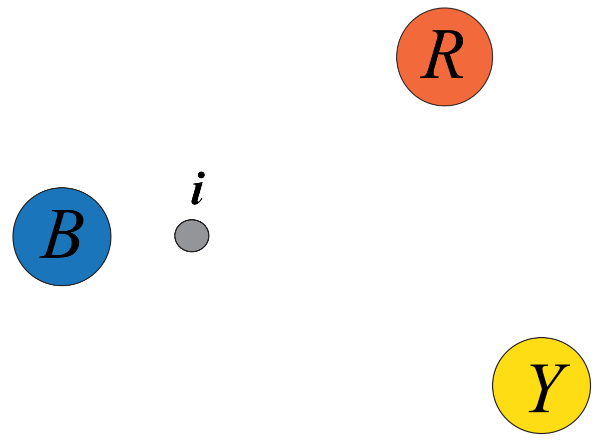
original space

2D space

$$C2 = KL(P\,||\,Q) = P\log\frac{P}{Q}$$

Similarity between instances and classes

Method



C2 = $KL(P \| Q)$

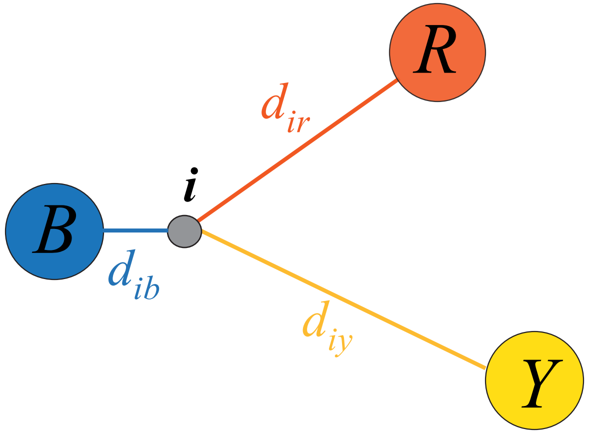
$\boldsymbol{P}$: Similarity between instances and classes in the original space

Class Probabilities

	B	R	Y
1	0.4	0	0.6
2	0.7	0.2	0.1
…	…	…	…
i	0.3	0.4	0.3
…	…	…	…

$\boldsymbol{P_i}$

Method



C2 = $KL(P \,||\, Q)$

Q : Similarity between instances and classes in the 2D space

Q_i

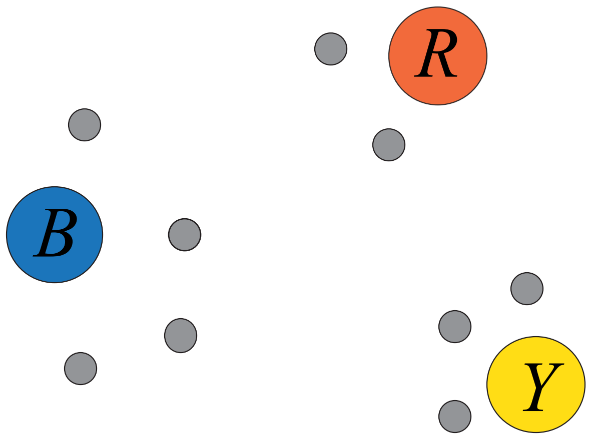
$$q_{ib} = \frac{(1 + d_{ib}^{2})^{-1}}{Sum}$$

$$q_{ir} = \frac{(1 + d_{ir}^{2})^{-1}}{Sum}$$

$$q_{iy} = \frac{(1 + d_{iy}^{2})^{-1}}{Sum}$$

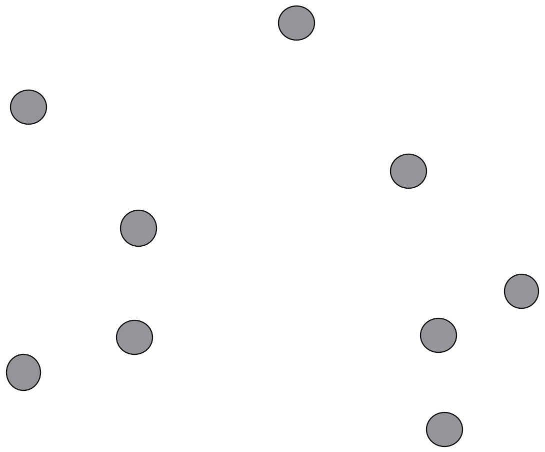
$$Sum = (1 + d_{ib}^{2})^{-1} + (1 + d_{ir}^{2})^{-1} + (1 + d_{iy}^{2})^{-1}$$

Method

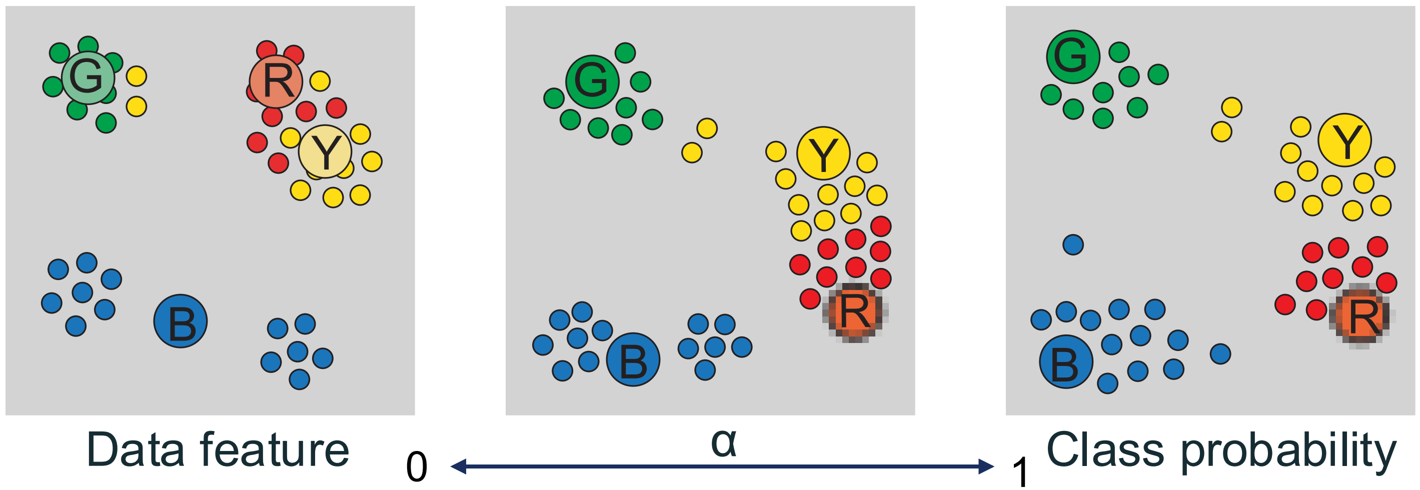


$$\text{C2} = \frac{1}{n}\sum_{i=1}^{n}\left(KL(P_i||Q_i) + \lambda \cdot D\right)$$

Method



Cost function: $C_d = (1-\alpha)\cdot \overbrace{C1}^{\text{Data feature structure}} + \alpha \cdot \underbrace{C2}_{\text{Class probability structure}}$

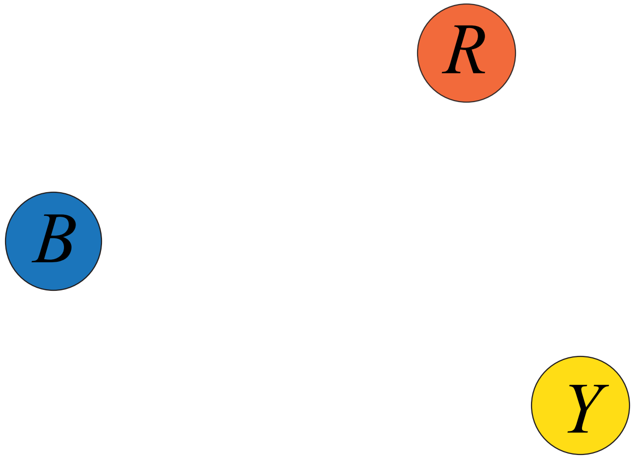


Data feature structure

Class probability structure

Method

Cost function:



$$C_c = C2$$

Class probability structure

Experiment

- Synthetic dataset
- Real-world datasets
  - Classifier Analysis
  - Document Topic Analysis

Comparative evaluation

- Use scenario
  - Visual interactive labeling

Fashion MNIST dataset

Class-constrained t-SNE

Load data

Load Data

Dataset
FashionMNIST

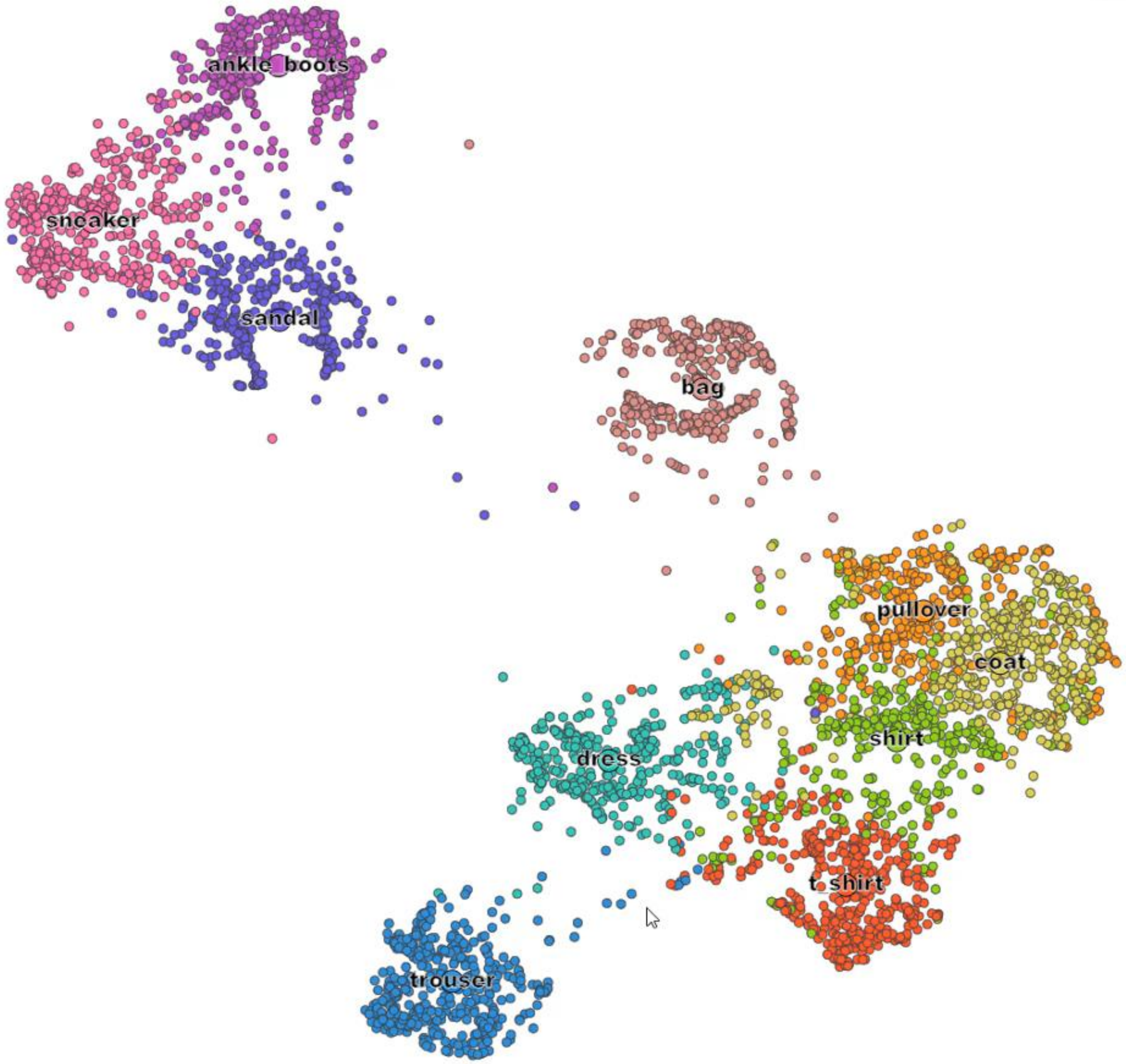
Model
CNN

Data Description: The Fashion MNIST dataset is composed of grayscale images of 28x28 pixels. 2000 images are selected, ensuring an equal distribution among ten classes, including T-shirts, trousers, pullovers, etc.

Adjust Parameters

α (structure balance parameter): 0.5

$\alpha = 0.5$

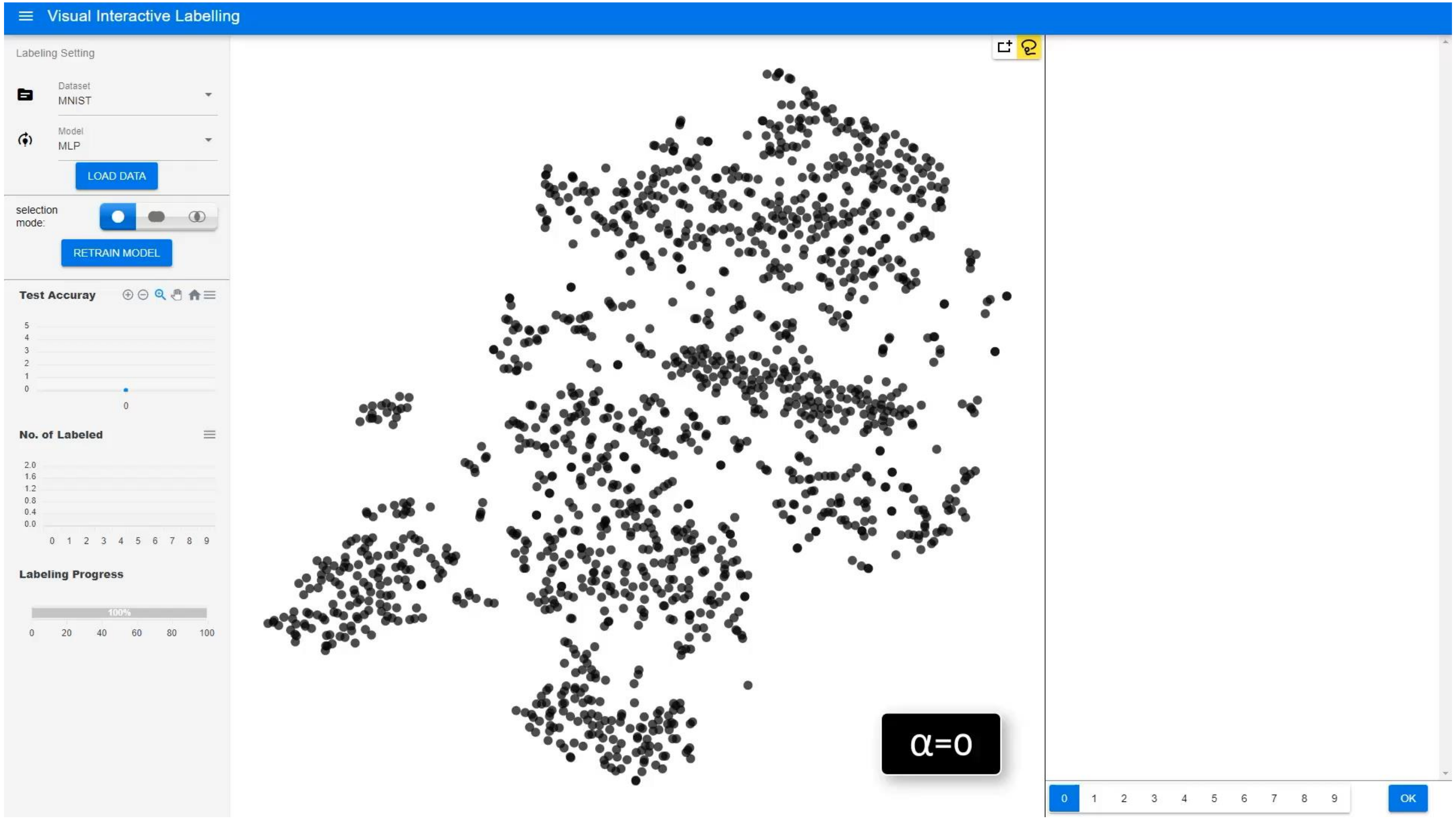


Use Case – Visual Interactive Labeling

Compact clusters/ equally spread

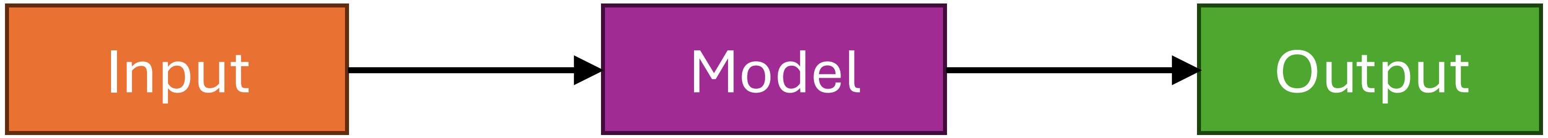
Coarse shapes of classes

Poorly separated classes/ outliers

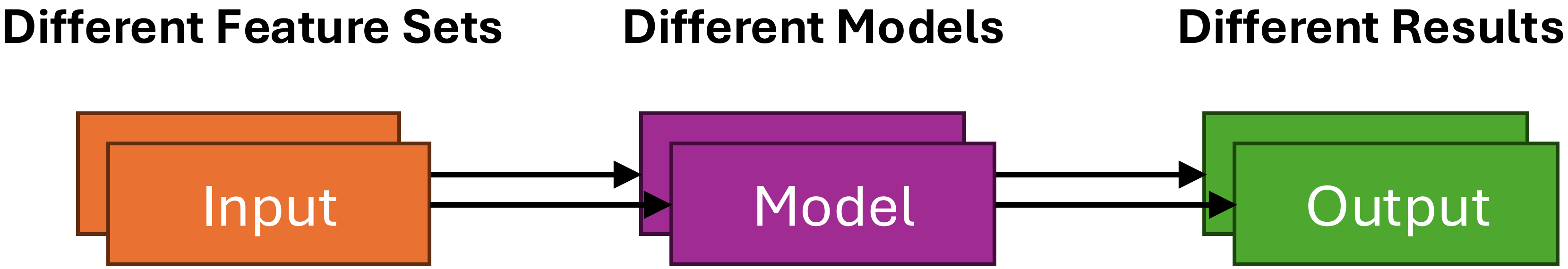


Bernard et al. 2018 Learning from the Best - Visual Analysis of a Quasi-Optimal Data Labeling Strategy

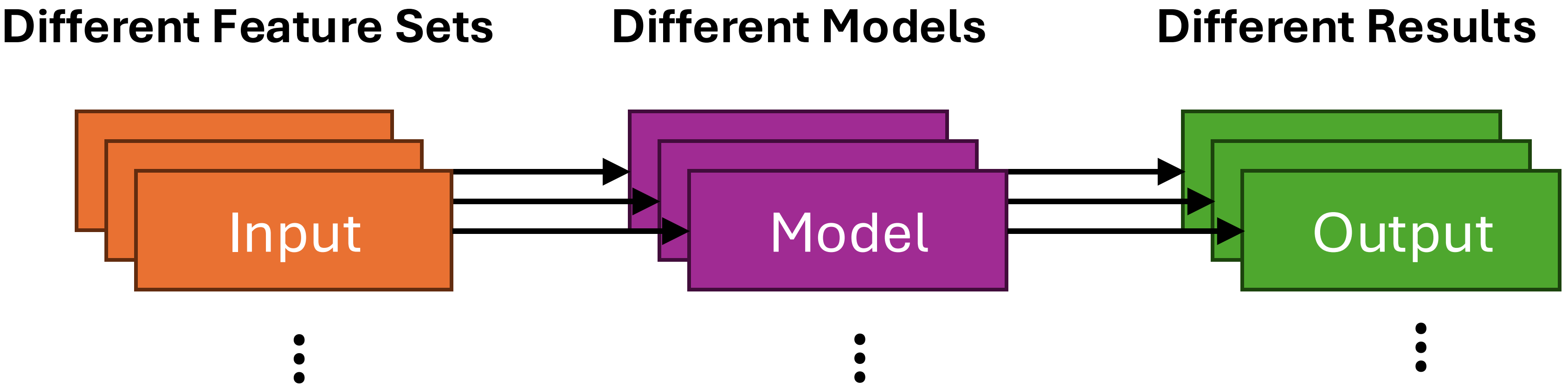
Main Perspectives in ML Workflow



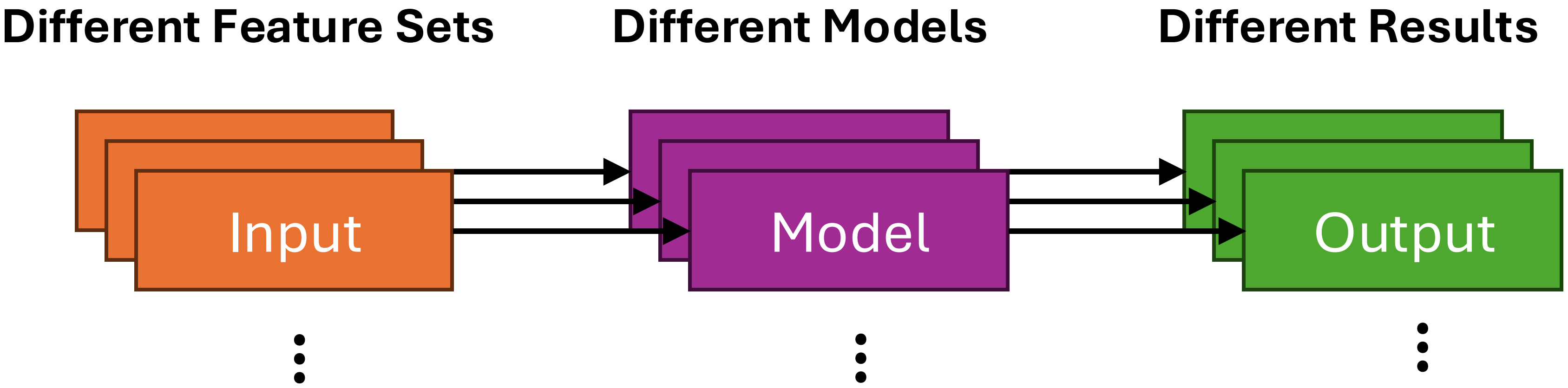
Main Perspectives in ML Workflow



Main Perspectives in ML Workflow

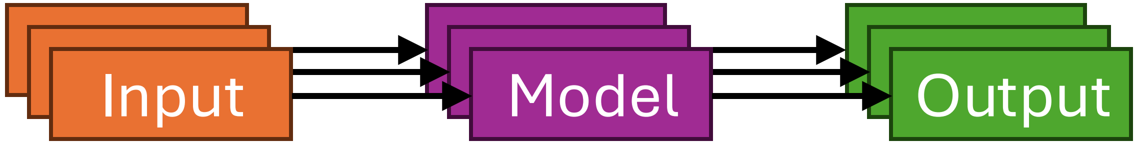


Main Perspectives in ML Workflow

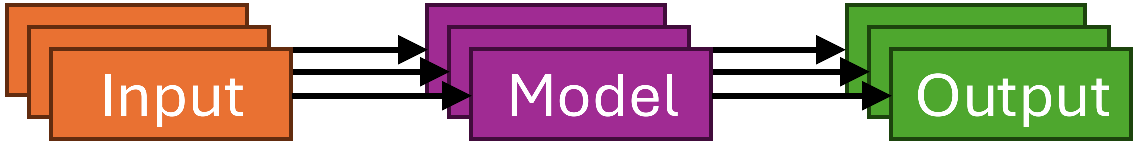


Model Comparison

Model Comparison

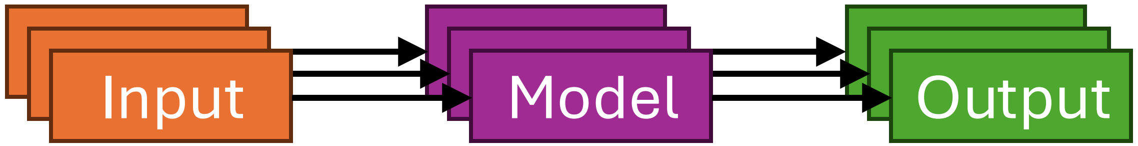


Model Comparison

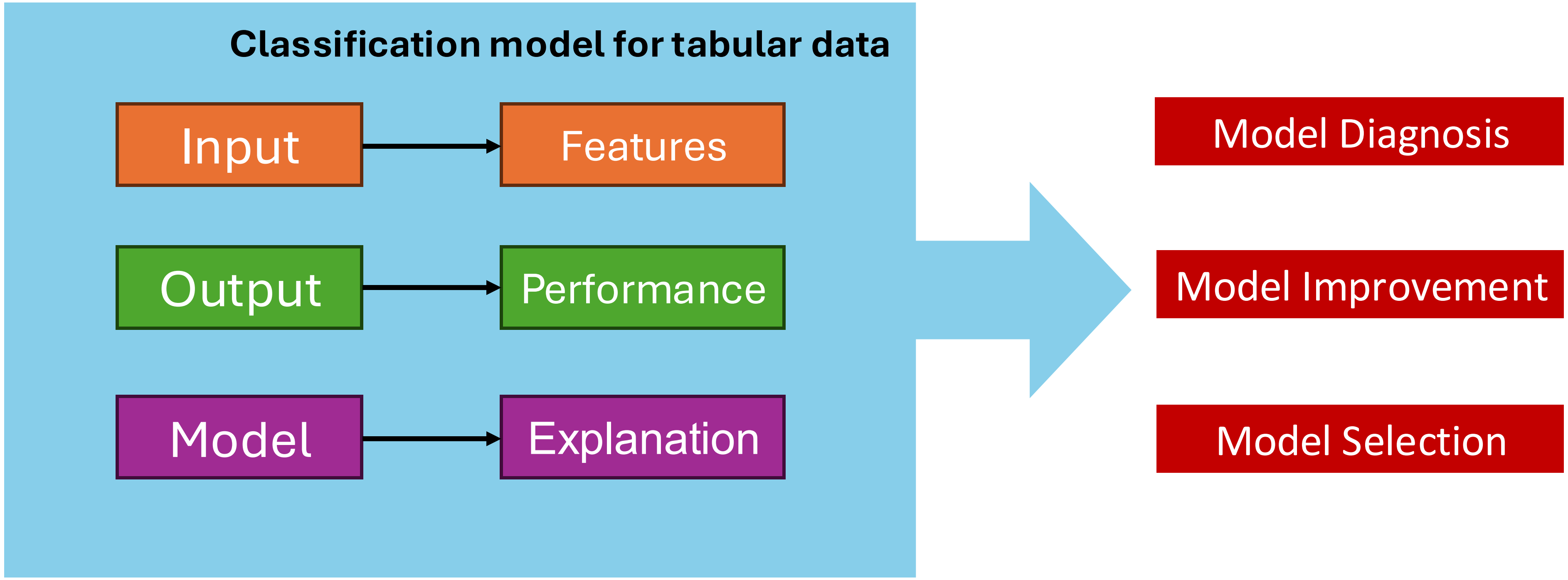


Input	Q: *Which features are used by different models?* Q: *Which features are most effective for the task?*	Model Diagnosis
Output	Q: *What are performance differences?* Q: *Where do models disagree with each other?*	Model Improvement
Model	Q: *Why do models (dis)agree with the classification?* Q: *Which model should I choose?*	Model Selection

ModelWise

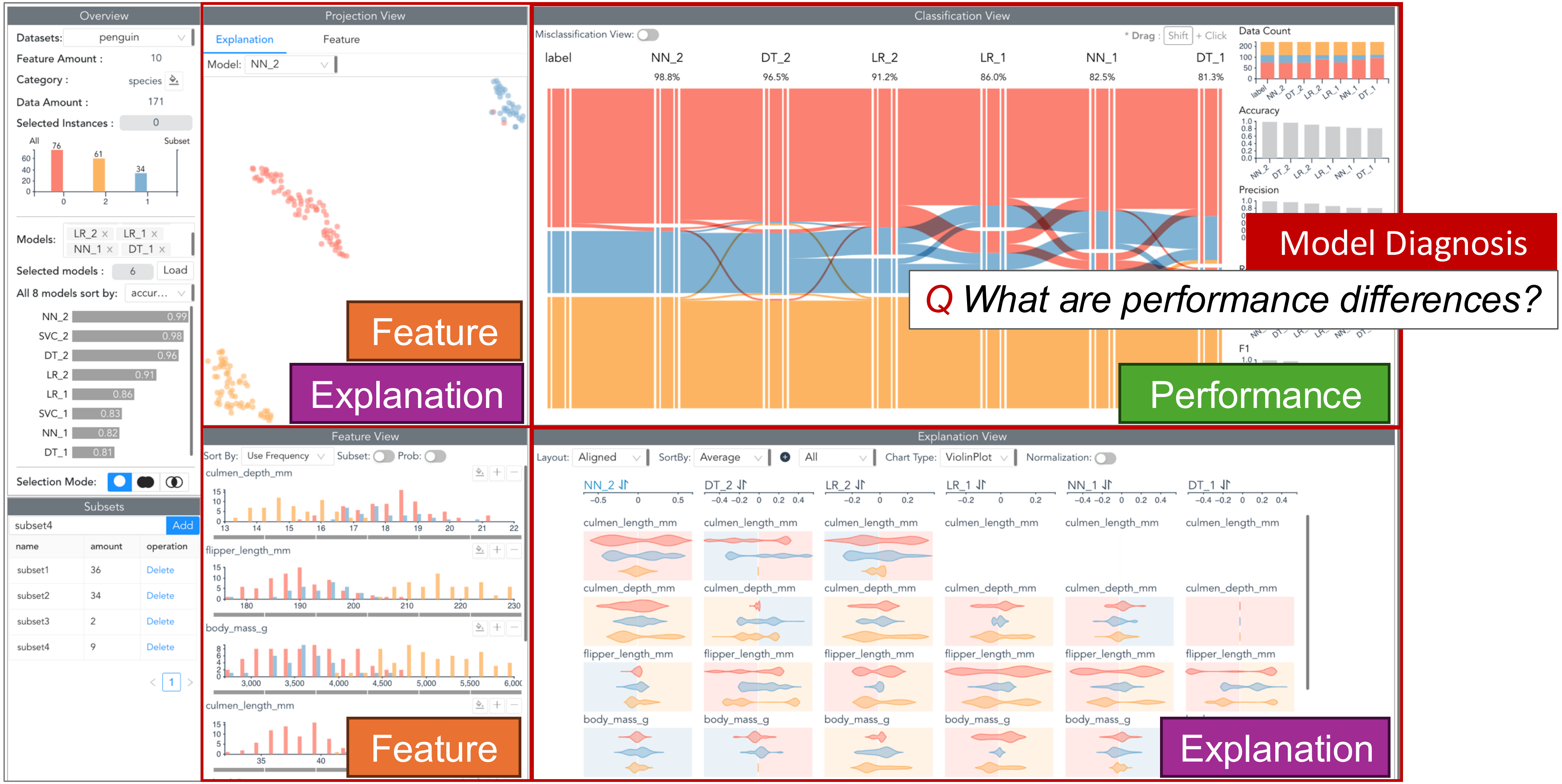


a visual analytics method to assist data scientists in comparing classification **model**s **wise**ly.



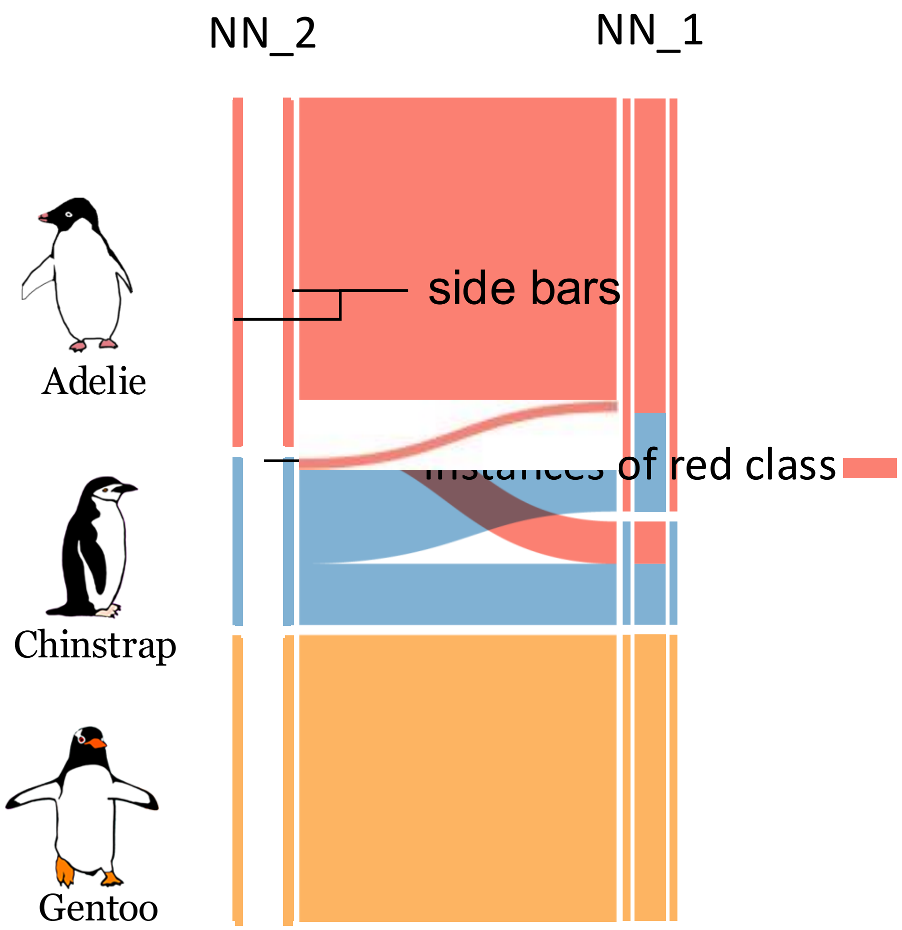
ModelWise

https://github.com/alicelh/ModelWise



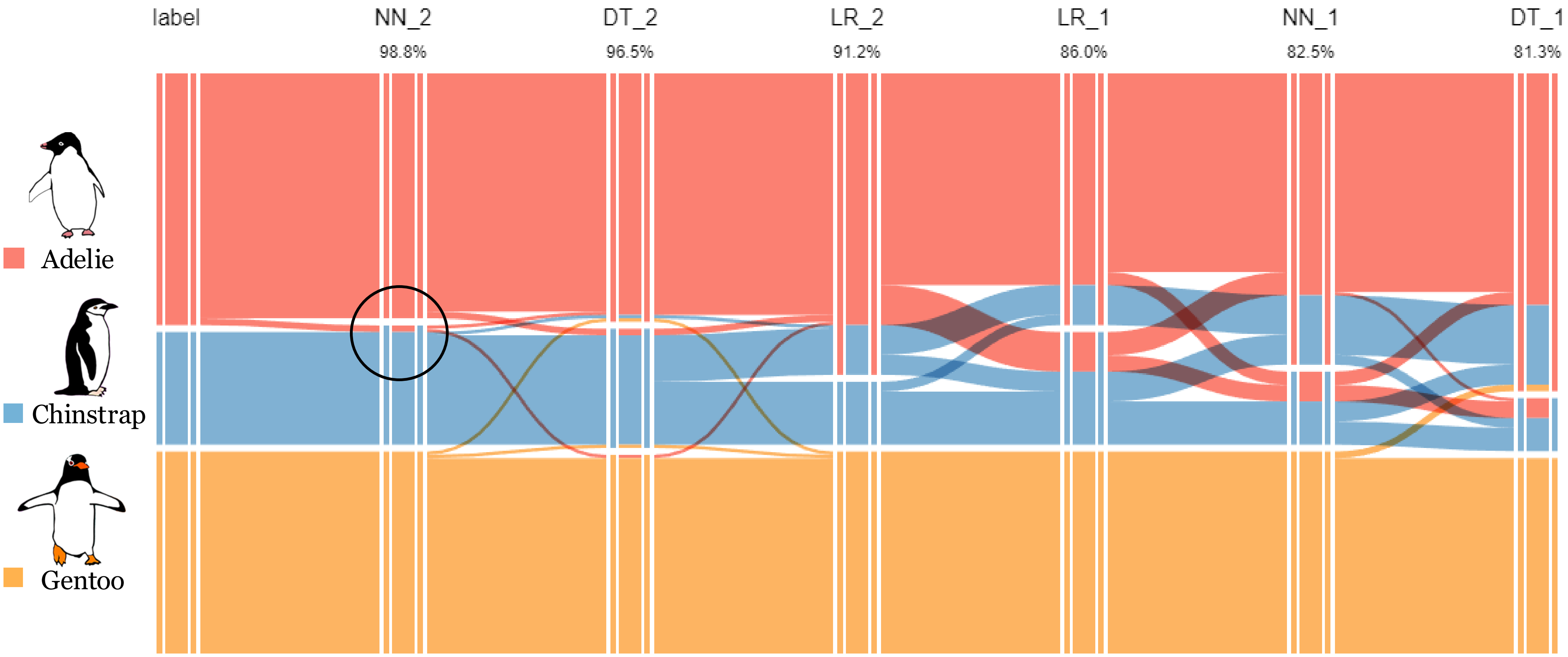
Performance

- Summary statistics
- Detailed performance



Performance

Confusion Sankey Diagram



Performance

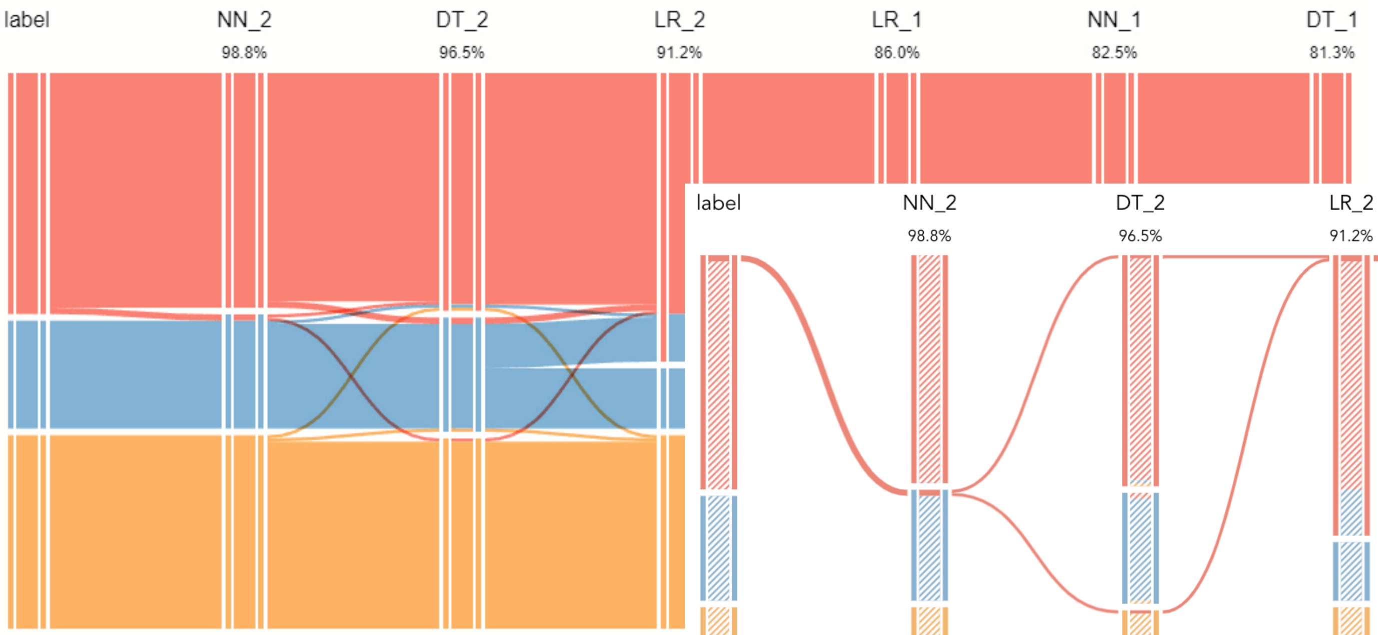
Confusion Sankey Diagram

Solution 1: Model Reordering

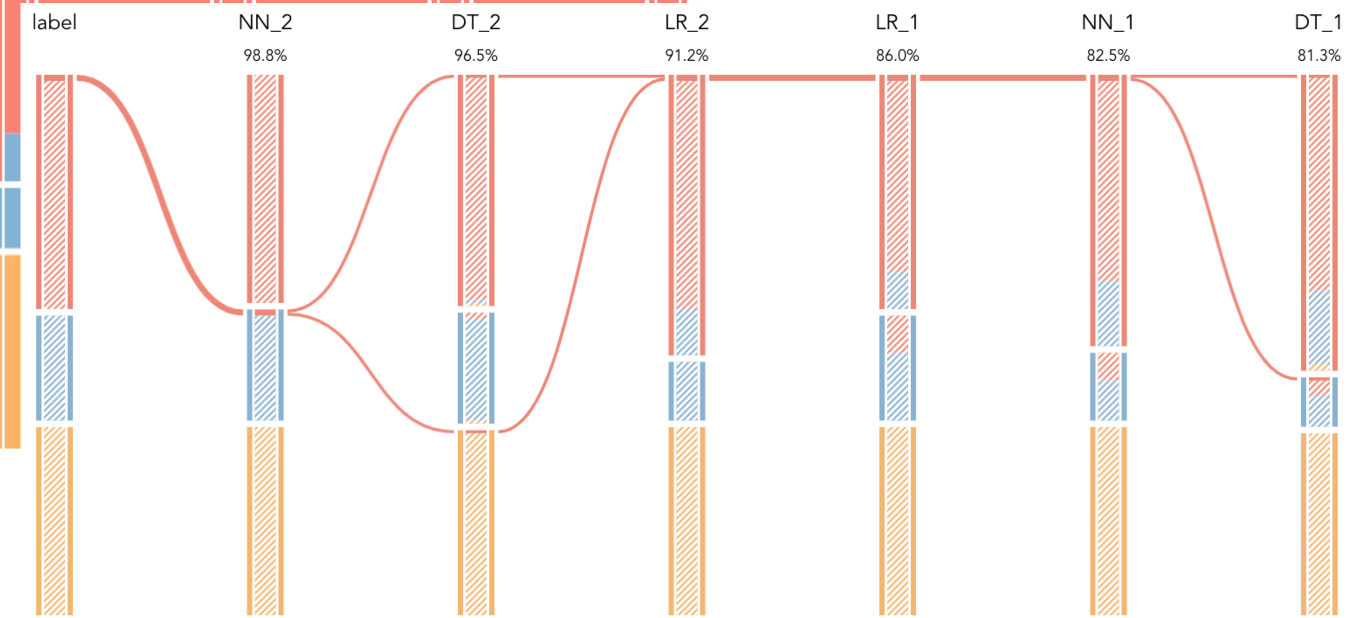
Adelie

Chinstrap

Gentoo

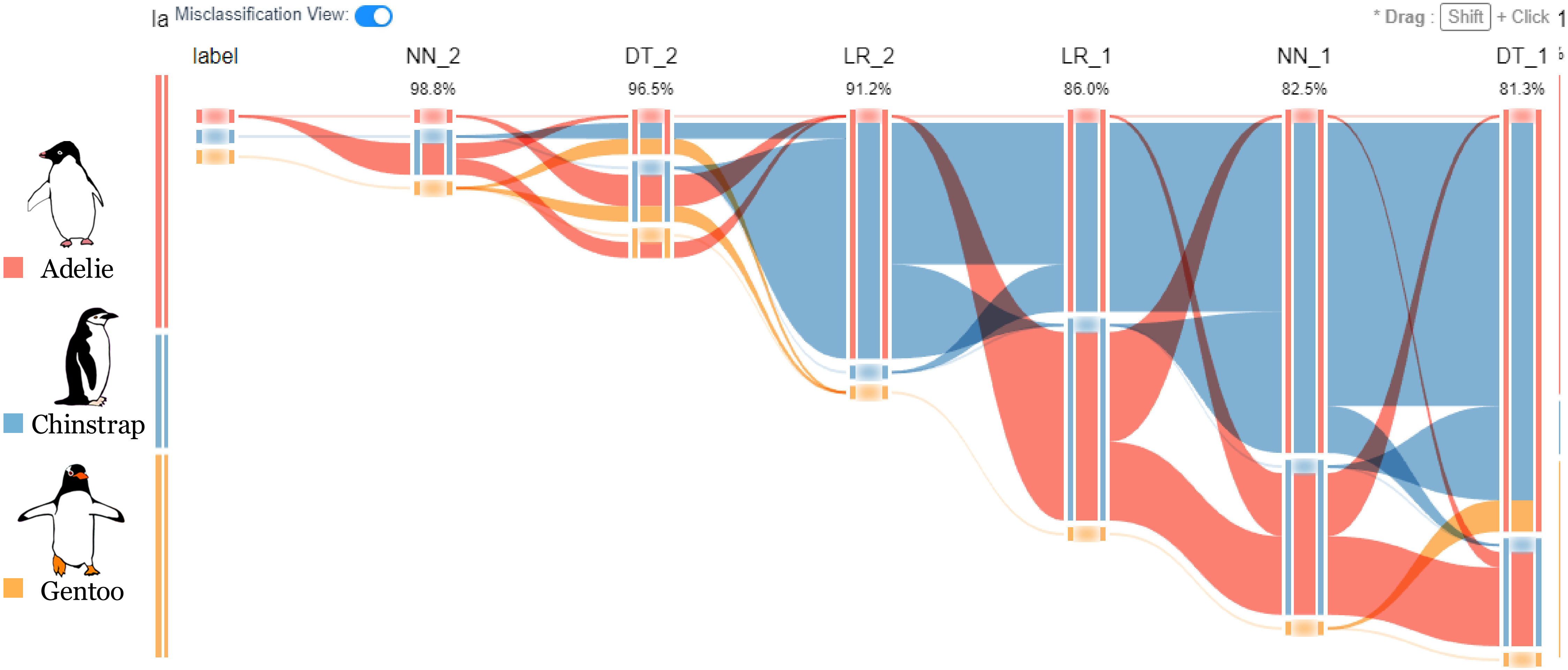


Solution 2: Subset Selection



Performance

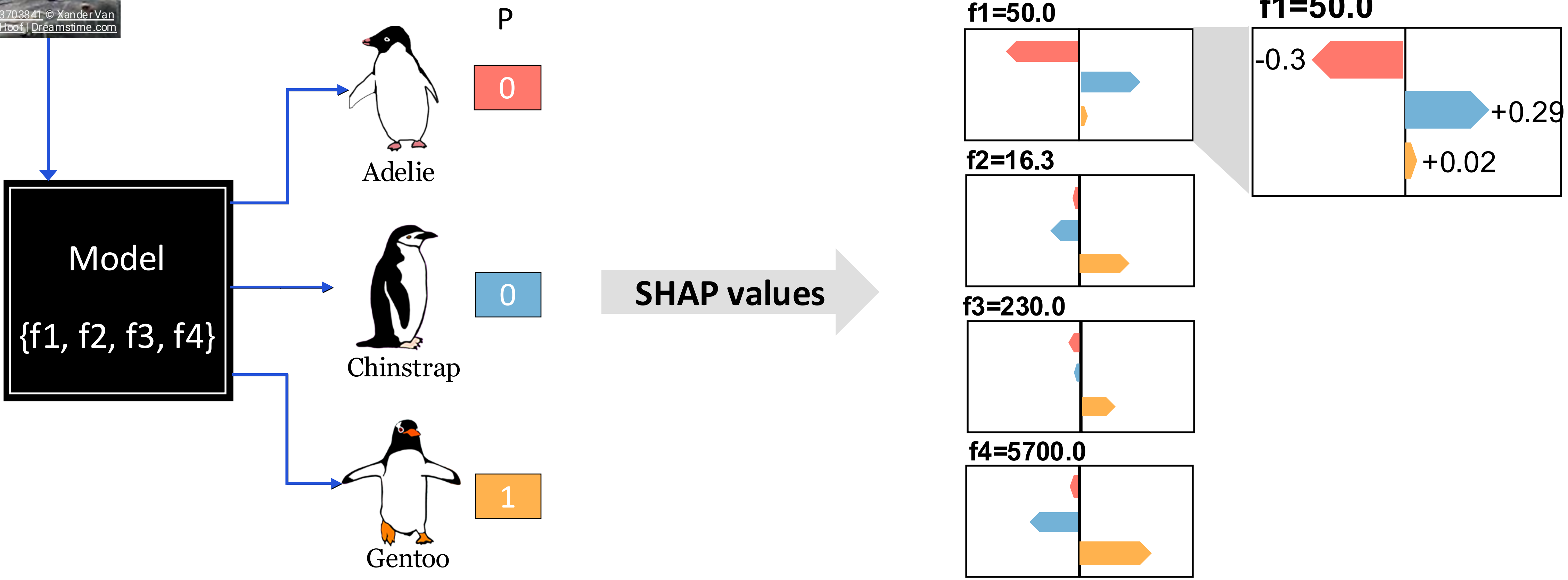
Confusion Sankey Diagram **Misclassification mode**



Explanation

Explanation

culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex	island		
					Biscoe	Dream	Torgersen
50.0	16.3	230.0	5700.0	0	1	0	0

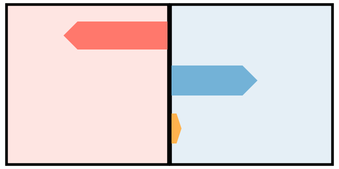


Explanation

- **Multiple feature sorting criteria**
- **Two layout methods**

One instance

f1(culmen_length_mm)

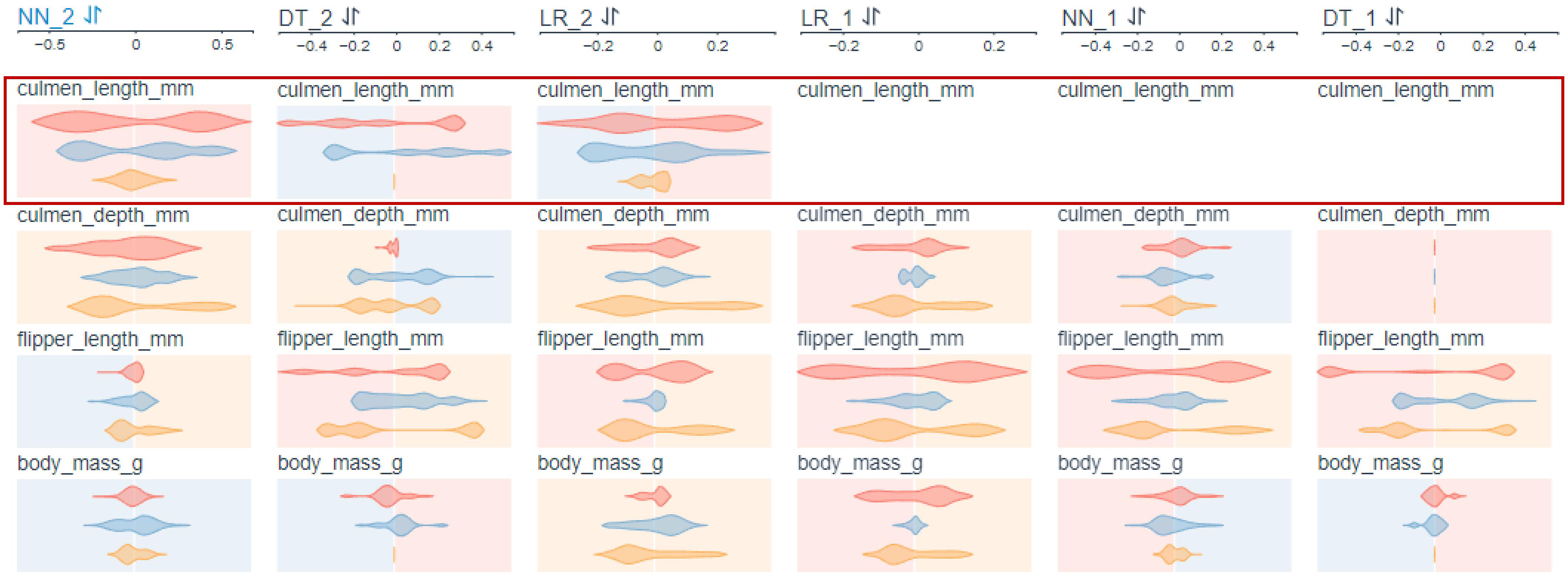


Multiple instances

Models

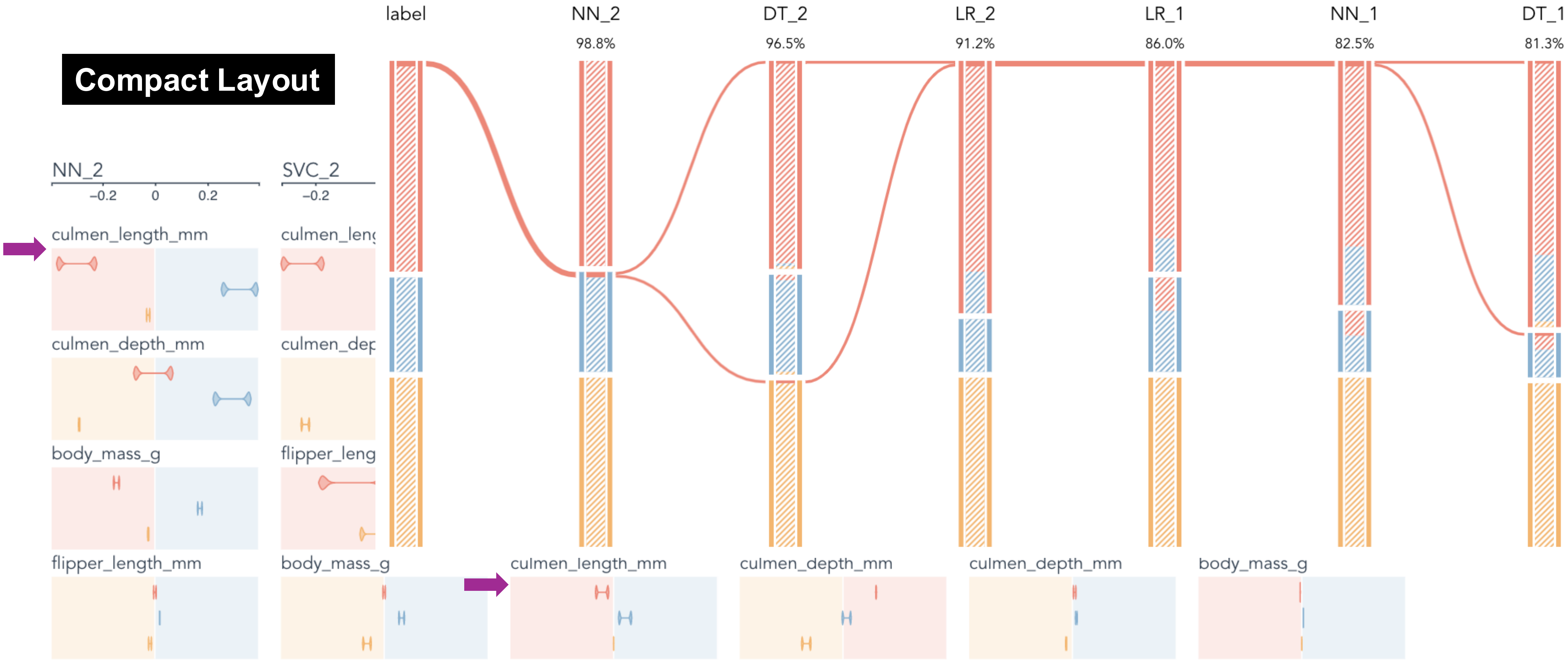
Aligned Layout

Features



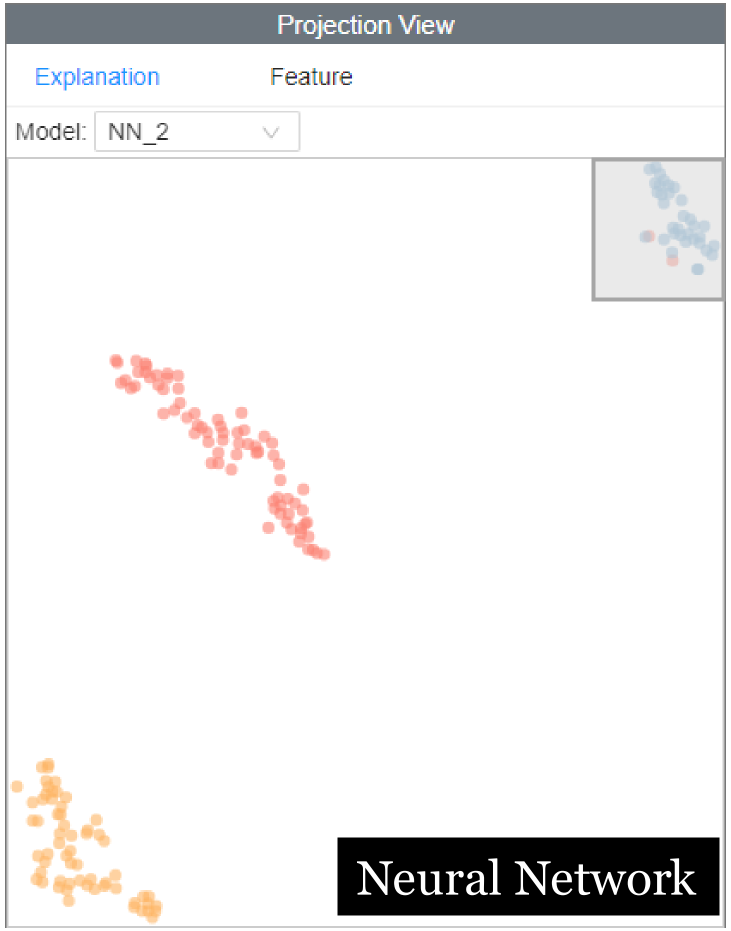
Explanation

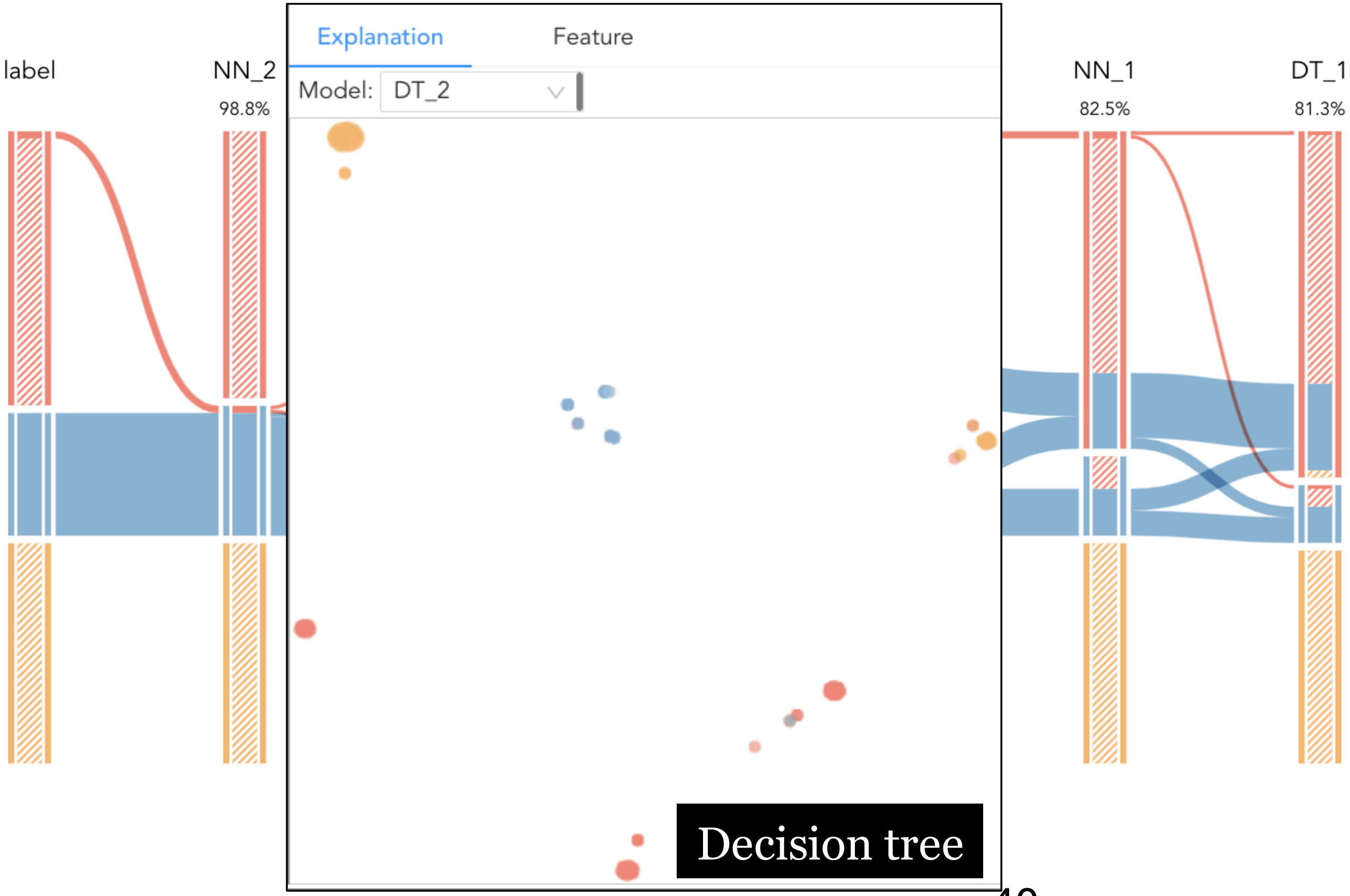
Compact Layout



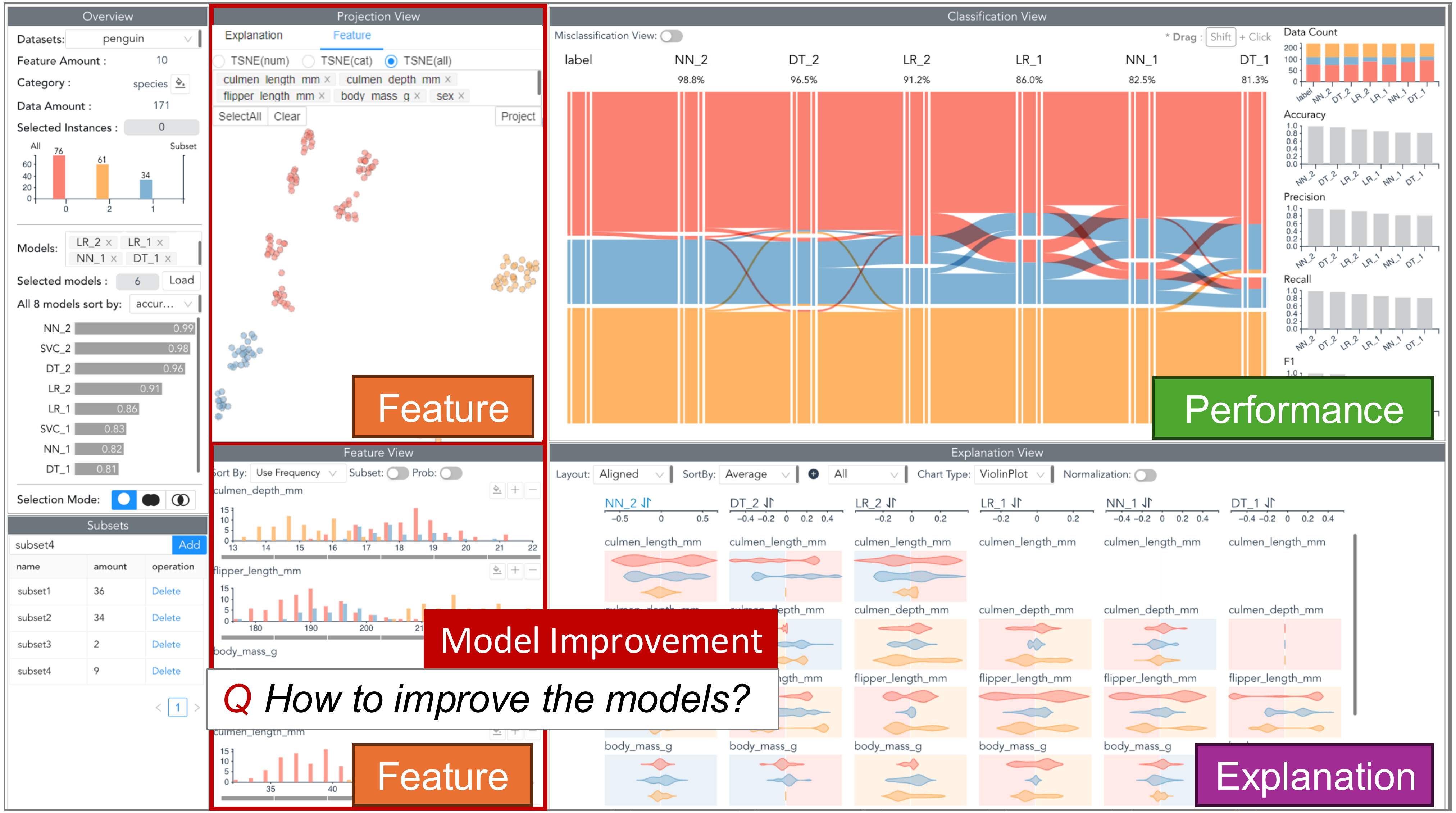
Explanation

Explanation Projection View





Feature


Overview
Projection View
Classification View
Feature View
Explanation View
Subsets
Feature
Feature
Performance
Model Improvement
Q How to improve the models?
Feature
Explanation

Case Study

Perioperative Deterioration Prediction

- Normal recovery (negative) versus potential unplanned ICU admission (positive)
- 44 variables
- Four models each with different algorithms and feature sets
  - Random Forest (RF),
  - Support Vector Machine (SVM)
  - Bayesian Network (BN)
  - Logistic Regression(LR)

Case Study

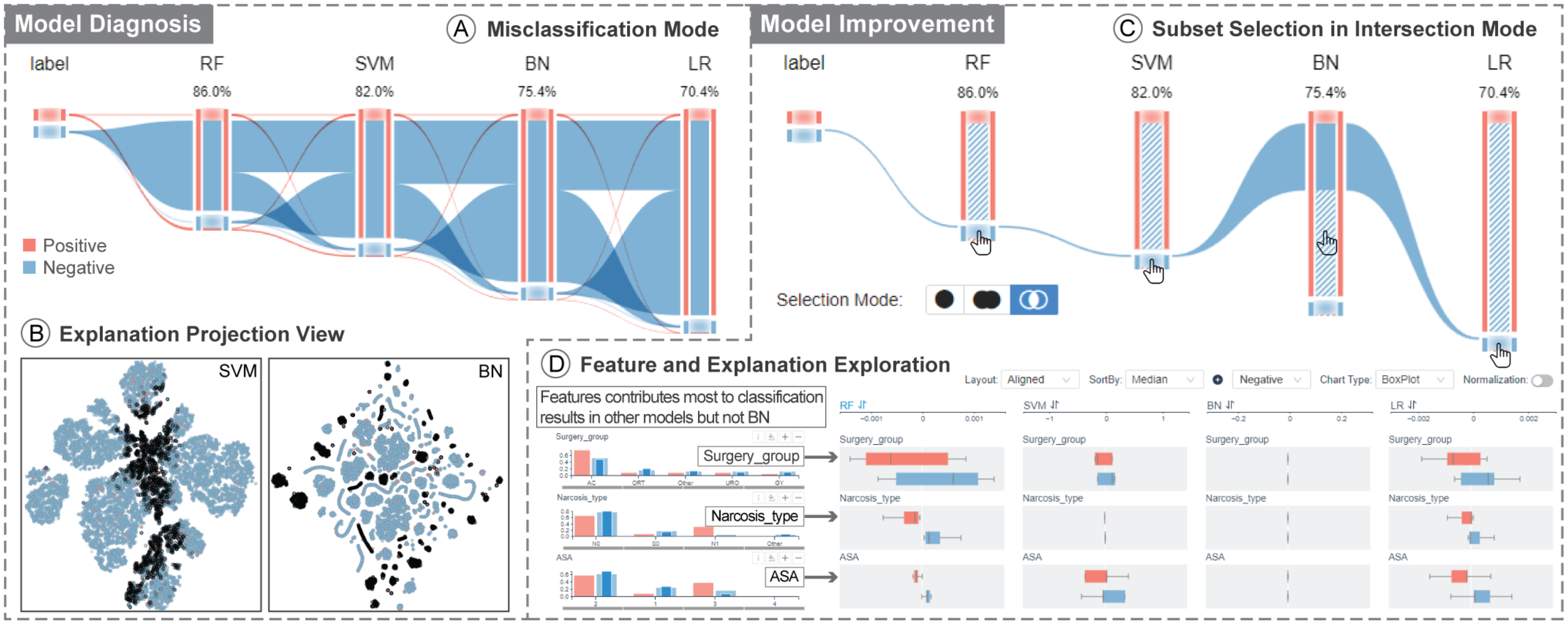
"I think it is a very nice way to explore what your model is doing. It gives you insight on model improvement, especially if you have different models to compare."

Model Diagnosis

- How does BN perform compared to other models?
- How does BN make classifications?

Model Improvement

- What can be learned from the other models to improve BN?



Who's behind SmartCHANGE?

We are a consortium of 14 international, multidisciplinary partners, with expertise in AI, healthcare, software engineering, social science and communication

Coordinator

Jožef Stefan Institute

COMMpla Communication Platforms and Online Solutions

connectedcare

ENGINEERING

jamk

U.PORTO UNIVERSIDADE DO PORTO

Università della Svizzera italiana

Univerza v Ljubljani

UNIVERSITY OF PIRAEUS RESEARCH CENTER

Trust-IT Services communicating to markets

TMU TAIPEI MEDICAL UNIVERSITY

TU/e

VUB VRIJE UNIVERSITEIT BRUSSEL

VUmc

Visit our website

Smart CHANGE

Empowering Youth: AI Models for a Healthier Future

smart-change.eu

smartchange-eu

@smartchange_eu

SmartChange-eu

Funded by the European Union

SmartCHANGE - Introduction

Joseph from Portugal

- 14 year old
- Does moderate and vigorous physical activity for 30 minutes everyday
- Eats two servings of red meat per week and 4 servings of whole grain cereals per day
- Sleeps 6 hours a day
- Plays online games 4 hours a day

Goal:

Estimate Joseph’s risk for cardiovascular diseases

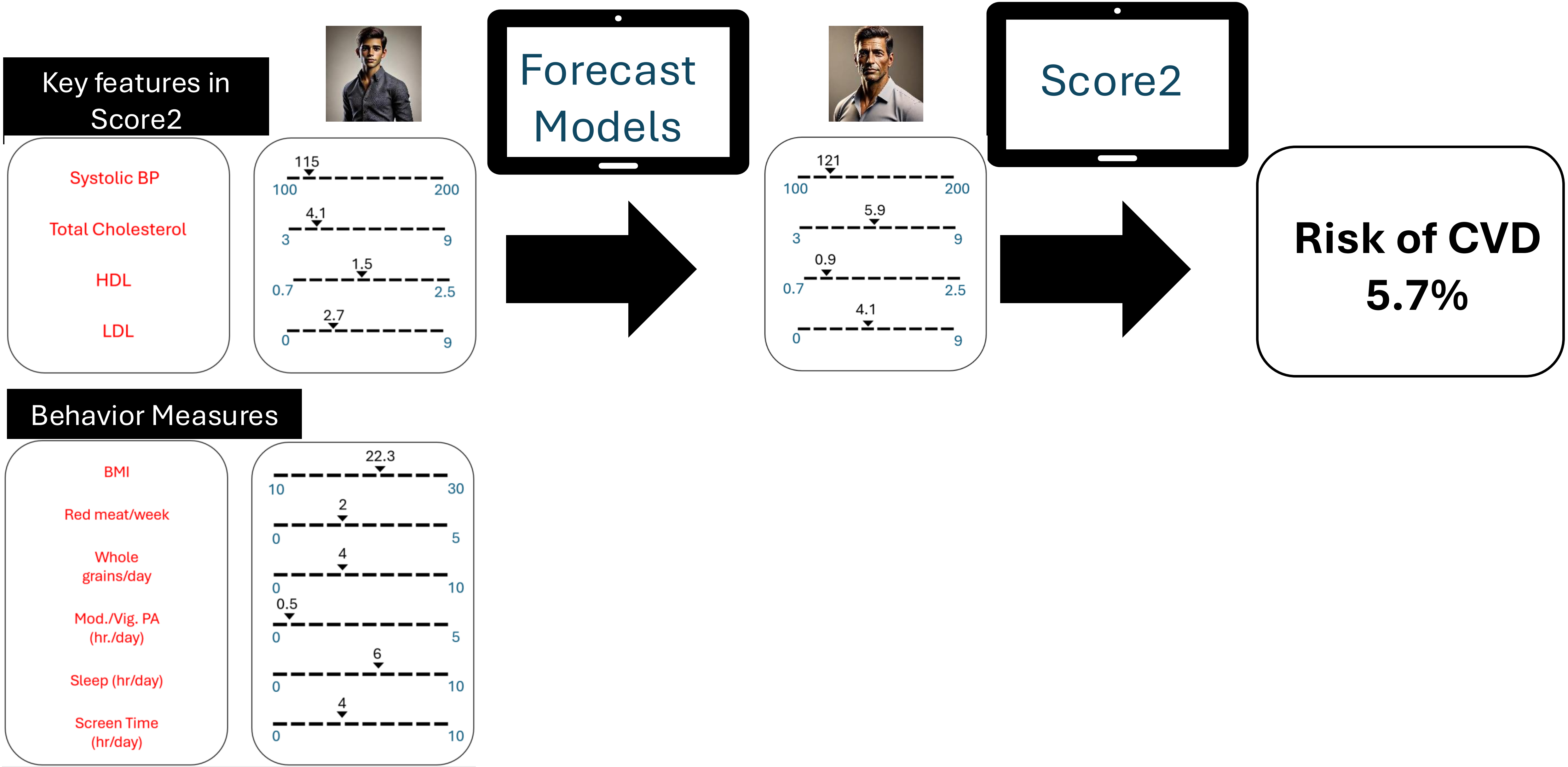
Score2

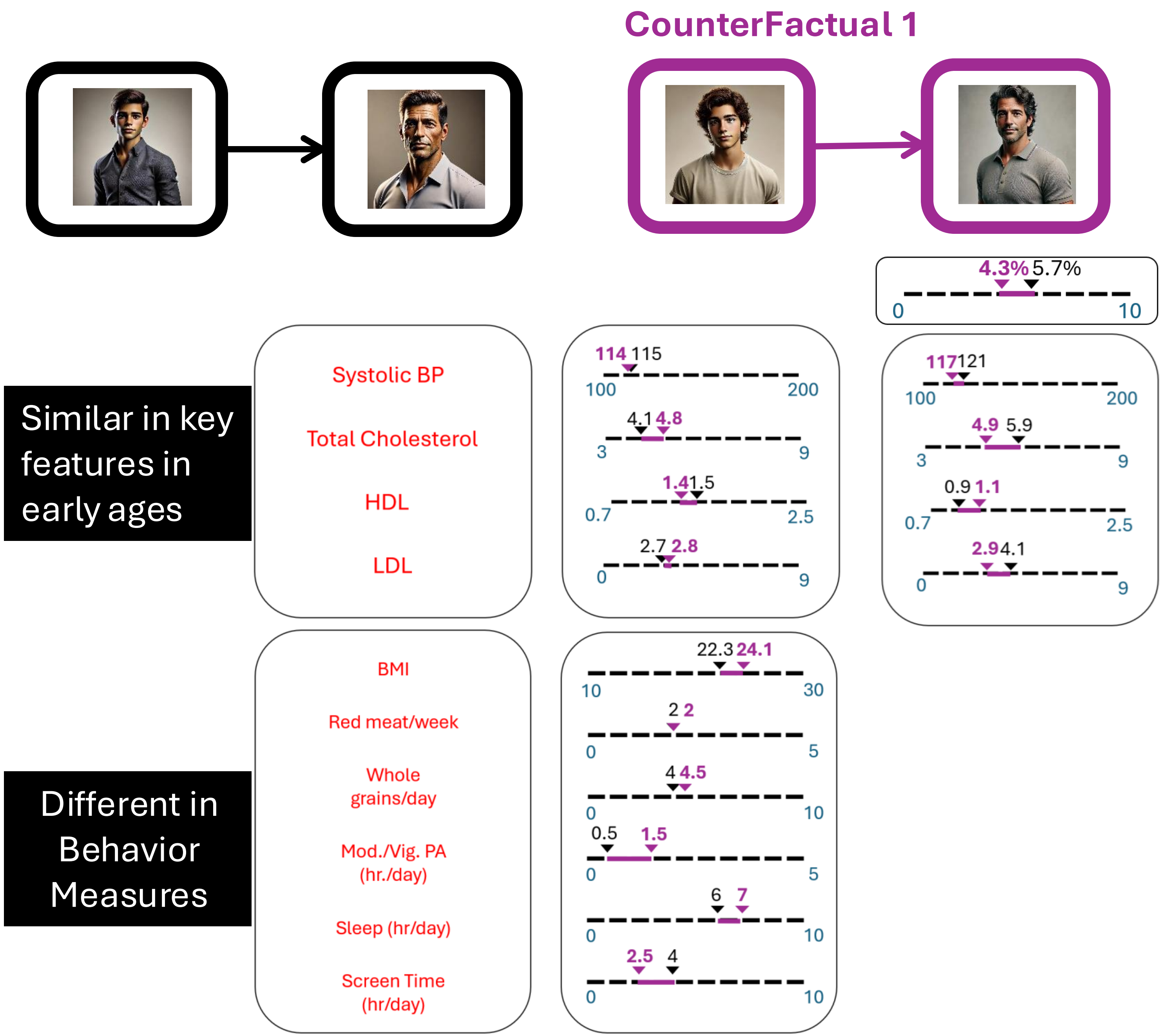
Features within user's control
Systolic Blood Pressure
Total Cholesterol
HDL
Smoker

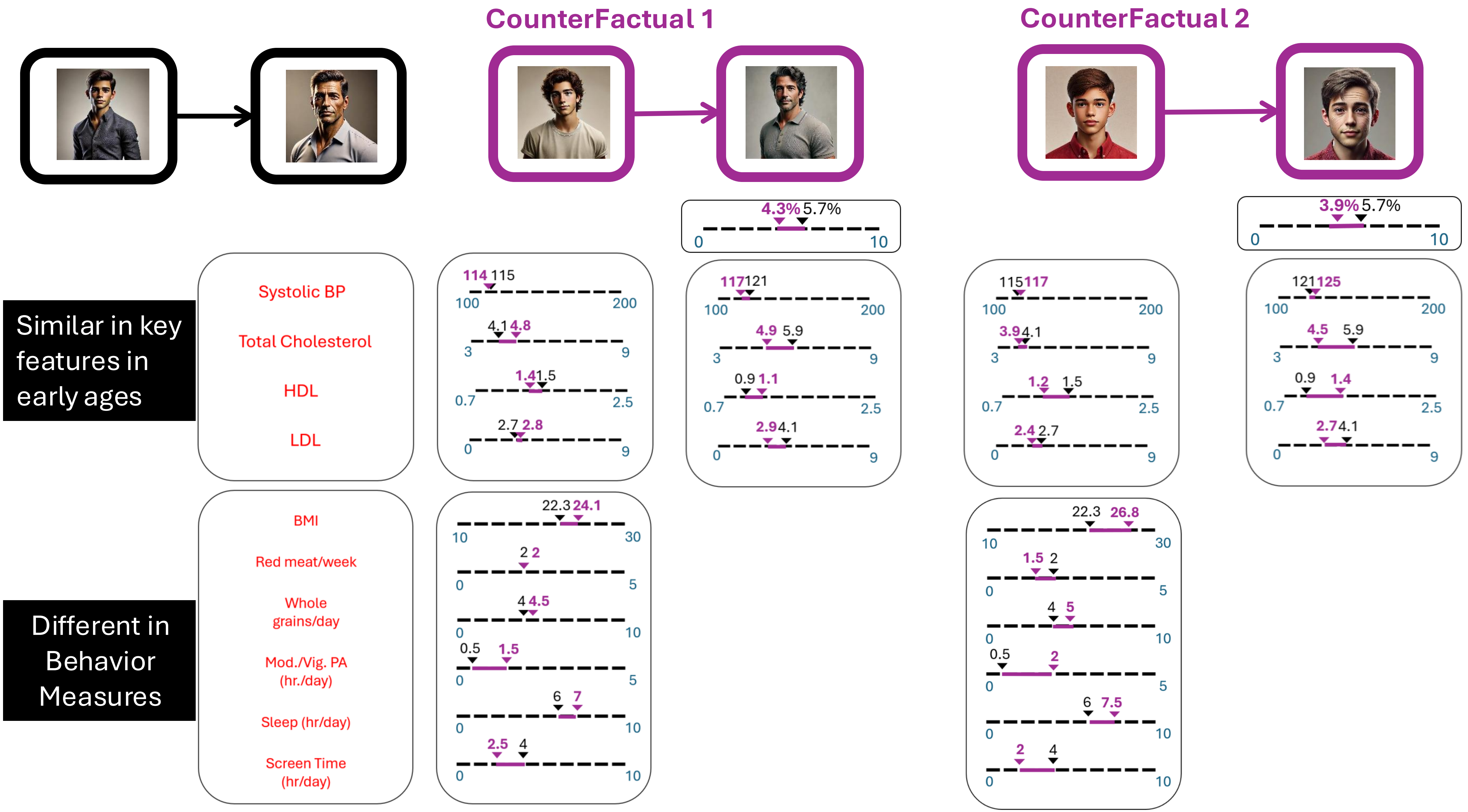
LDL (used for recommendations)

Features not within user's control
Age
Sex
Diabetes History
Risk Region

What does SmartCHANGE do?

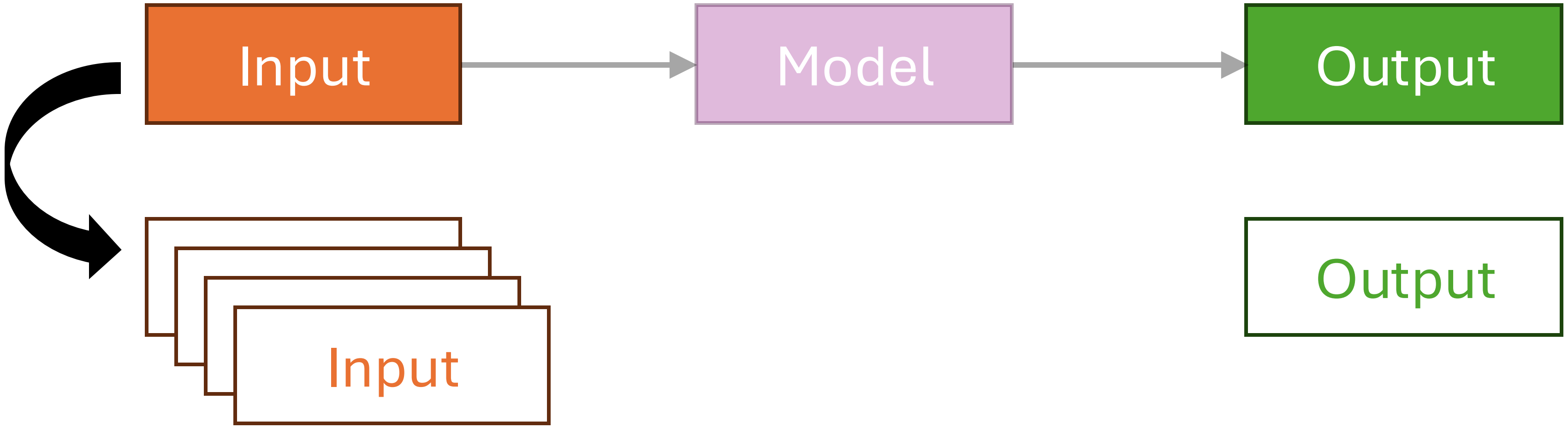



CounterFactual 1
4.3% 5.7%
0
10
Similar in key features in early ages
Systolic BP
Total Cholesterol
HDL
LDL
114 115
100
200
4.1 4.8
3
9
1.4 1.5
0.7
2.5
2.7 2.8
0
9
117 121
100
200
4.9 5.9
3
9
0.9 1.1
0.7
2.5
2.9 4.1
0
9
Different in Behavior Measures
BMI
Red meat/week
Whole grains/day
Mod./Vig. PA (hr./day)
Sleep (hr/day)
Screen Time (hr/day)
22.3 24.1
10
30
2 2
0
5
4 4.5
0
10
0.5 1.5
0
5
6 7
0
10
2.5 4
0
10

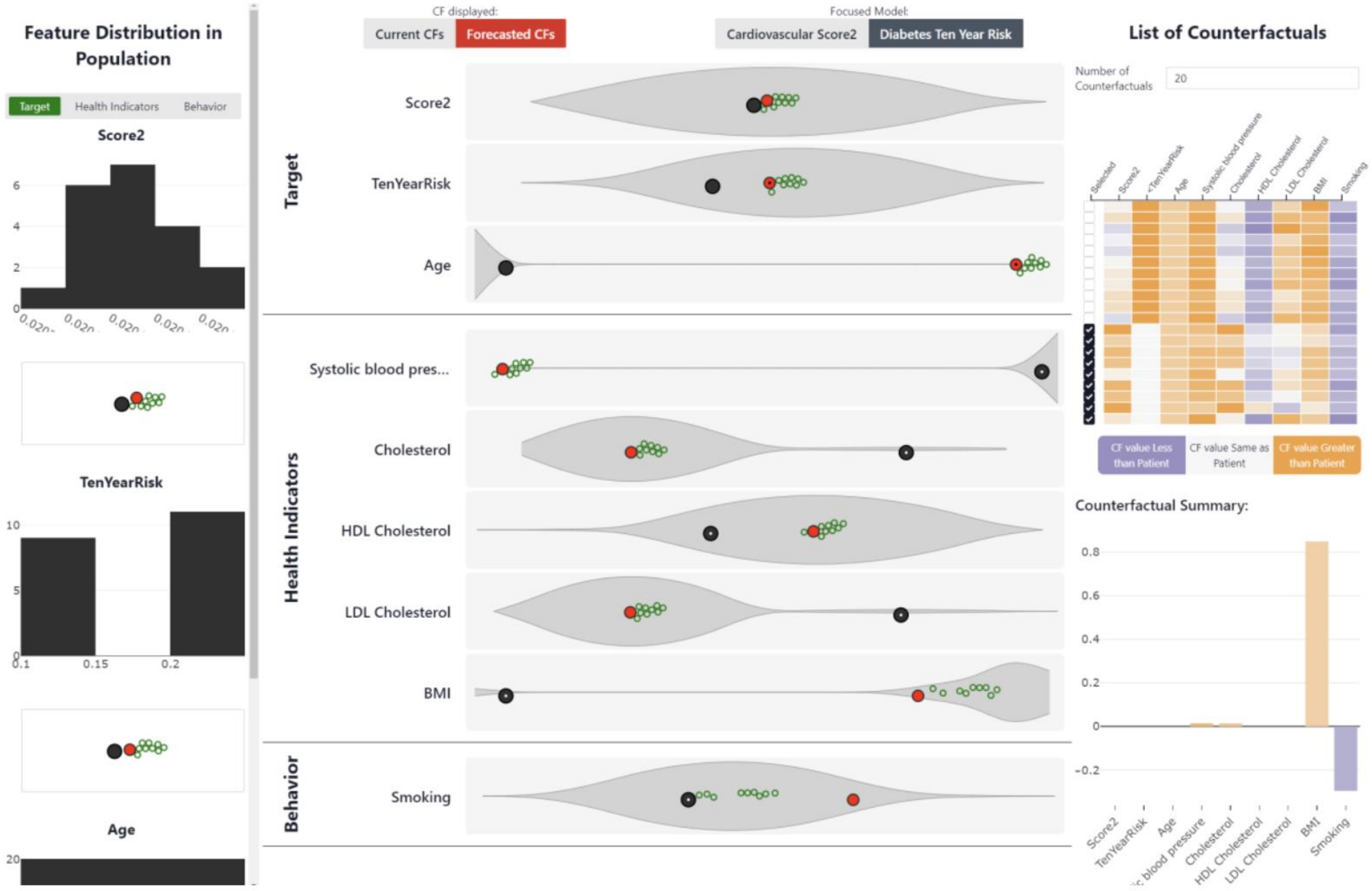


CounterFactual 1
CounterFactual 2
Similar in key features in early ages
Different in Behavior Measures
Systolic BP
Total Cholesterol
HDL
LDL
BMI
Red meat/week
Whole grains/day
Mod./Vig. PA (hr./day)
Sleep (hr/day)
Screen Time (hr/day)
4.3% 5.7%
3.9% 5.7%
114 115
4.1 4.8
1.4 1.5
2.7 2.8
117 121
4.9 5.9
0.9 1.1
2.9 4.1
115 117
3.9 4.1
1.2 1.5
2.4 2.7
121 125
4.5 5.9
0.9 1.4
2.7 4.1
22.3 24.1
2 2
4 4.5
0.5 1.5
6 7
2.5 4
22.3 26.8
1.5 2
4 5
0.5 2
6 7.5
2 4

Counterfactual Explanations



Visual explainer



Thanks for listening!

Eindhoven University of Technology
Visualization cluster

Linhao Meng

Stef van den Elzen

Anna Vilanova