

# Explainable AI in Credit Risk Management

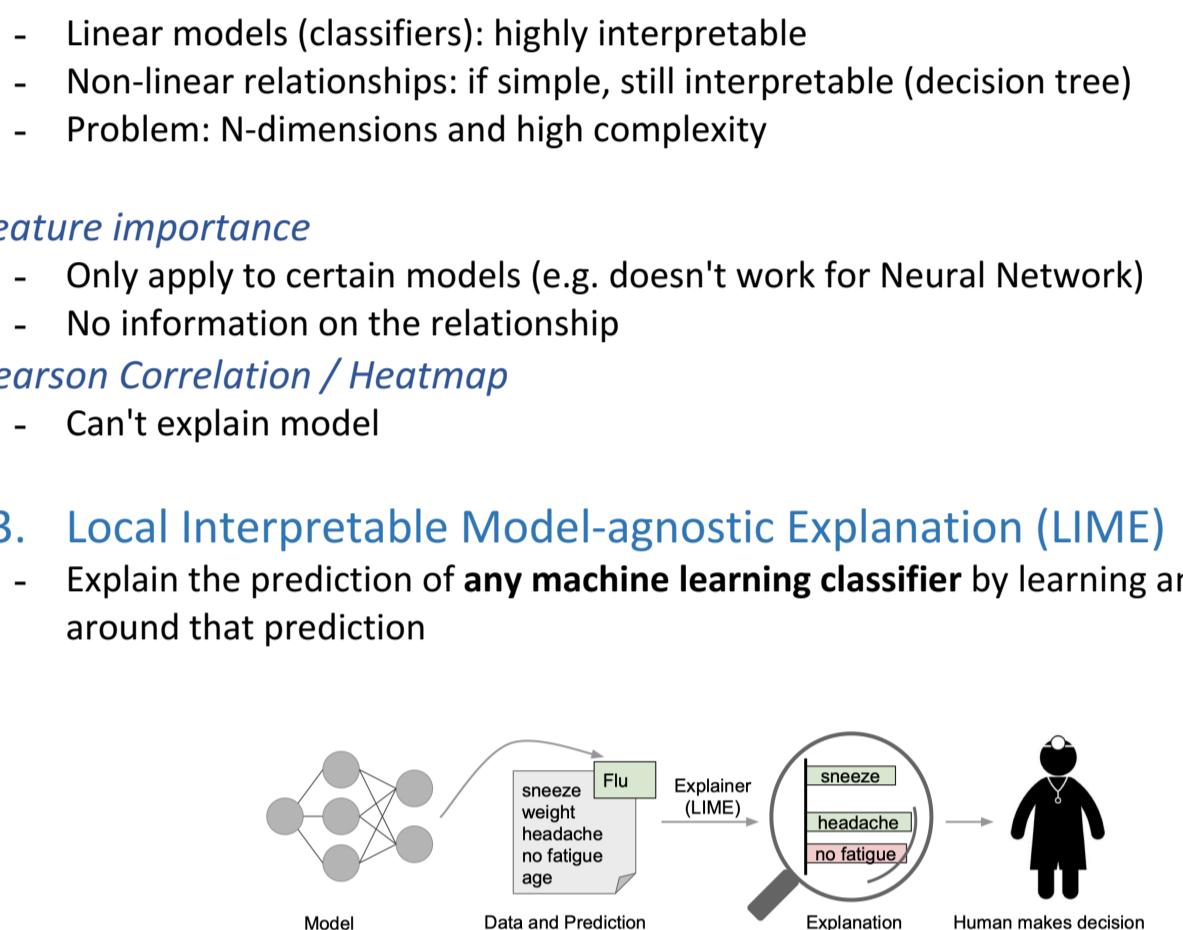
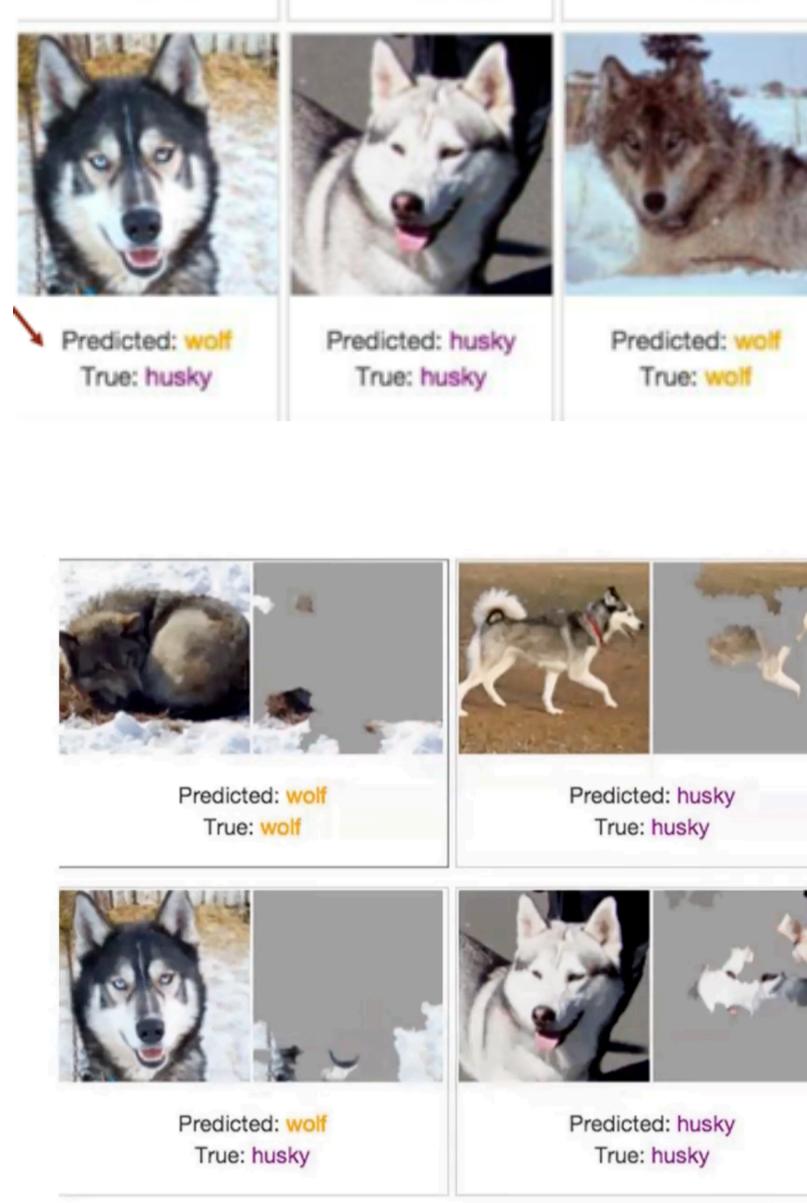
*Speaker: Branka M.*  
Date: 09/17/2021

## 1. Why do we need interpretability?

- Predictions are made from wrong information

A collage of three photographs showing different dogs. The first image is a close-up of a dog's head with dark brown and white fur. The second image shows a dog with a white and grey coat running across a grassy field. The third image is a close-up of a dog's face, possibly a Husky, with a mix of brown, black, and white fur.

- Predicted: wolf      Predicted: husky      Predicted: wolf



works

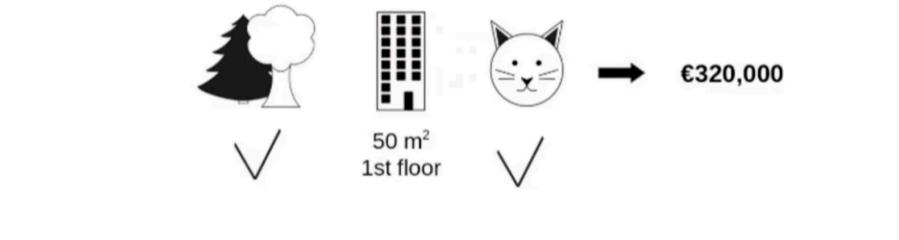
- $g \in G$ : An explanation considered as a model
  - $f: \mathbb{R}^d \rightarrow \mathbb{R}$ : The main classifier being explained
  - $\pi_x(z)$ : The proximity measure of an instance  $z$  from  $x$
  - $\Omega(g)$  - Complexity parameter (e.g. number of features)

- $L$  is the median since a key

- ## 1 Shapley Values

- The Shapley value is the average marginal contribution of a feature value

possible coalitions.



Given a model  $f(x_1, x_2, x_3 \dots x_n)$  with feature 1 to  $n$  being payers in a game in which the payoff  $v$  is the measure of importance of the subset.

- Let  $\Pi$  be the set of permutations of the integers up to  $N$ , and given  $\pi \in \Pi$ ,  $\pi(i)$  are the players preceding player  $i$  in  $\pi$ , then:

$$N! \sum_{\pi \in \Pi}$$

Our data points should have similar outputs and similar explanations.

## Stability of Explanation through Graph Theory

Create a fully connected graph for each data

Create a fully connected graph for each data points  
distance

Minimum Spanning Tree (MST)

**Minimum Spanning Tree (MST)**  
Find a spanning tree that connect all nodes with minimum total weight.

- ```
Top Features k=0
Green 25 : open_il_12m          Count : 67
Blue   2 : grade                Count : 2
Grey   27 : mths_since_rcnt_il Count : 32
```



- Pick random data point, compare top 5 important features both by LIME and Shapley  
Result is consistent

| Prediction probabilities | Fully Paid | Default                | Feature        |
|--------------------------|------------|------------------------|----------------|
| Fully Paid               | 0.77       | inq_last_6mths <= 0.00 | inq_last_6mths |

Fully Paid   
Default 

