www.pmu.edu.sa

MassiveAnalytic
Precognition Machines
www.massiveanalytic.com

Non-specificity-based Supervised Discretization for Possibilistic Classification

Ilyes Jenhani

Prince Mohammad Bin Fahd University, Khobar, KSA

Massive Analytic Ltd., London, UK

Ghaith Khlifi

Massive Analytic Ltd., London, UK

Panagiotis Sidiropoulos

Massive Analytic Ltd., London, UK

Mullard Space Science Laboratory, University College London, UK

Henk Jansen

Massive Analytic Ltd., London, UK

George Frangou

Massive Analytic Ltd., London, UK

UCL Quantum Science & Technology Institute, UK

Autonomous Systems Dynamics and Control Research Group Centre for Autonomous and Cyber Physical Systems, Cranfield University, UK

October 19, 2022

Context and Motivation

MAL - Decision Intelligence

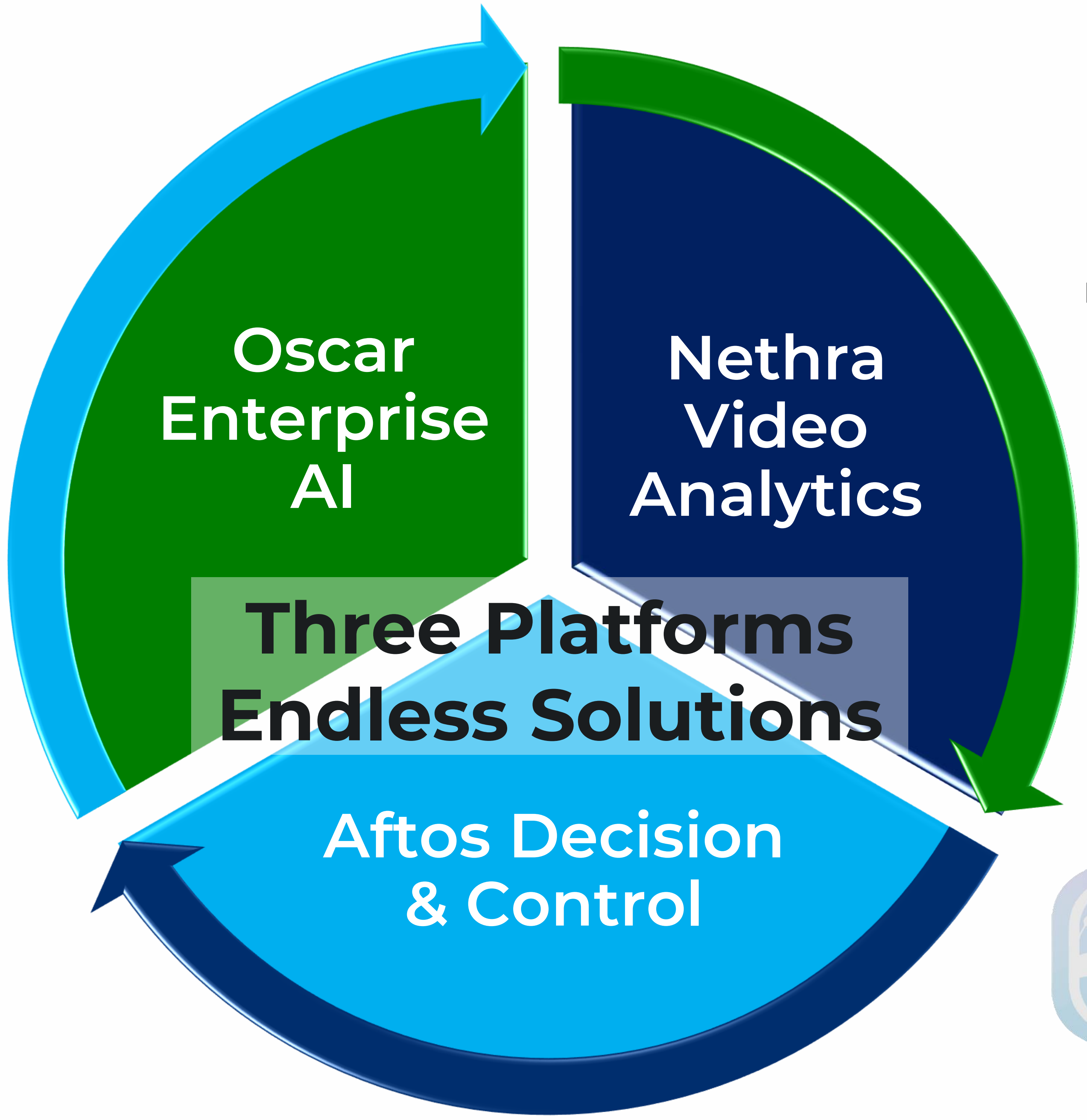
Massive Analytic is a Deep-tech Scale-up specialising in Decision Intelligence.

MAL's suite of platforms can be integrated for implementation to generate world-beating decision support and control outputs.

Combine and Analyse

- Data Fusion
- Data Processing
- Analysis and learning
- Create models

OSCAR
DATA SCIENCE



NETHRA
VIDEO ANALYTICS

Generate Data from video and imagery

- Extract information from video imagery
- Objects, movement and behaviour
- Flexible models and complex workflows

AFTOS
ROBOTICS

Automate, Simulate, Predict, Decide and Control

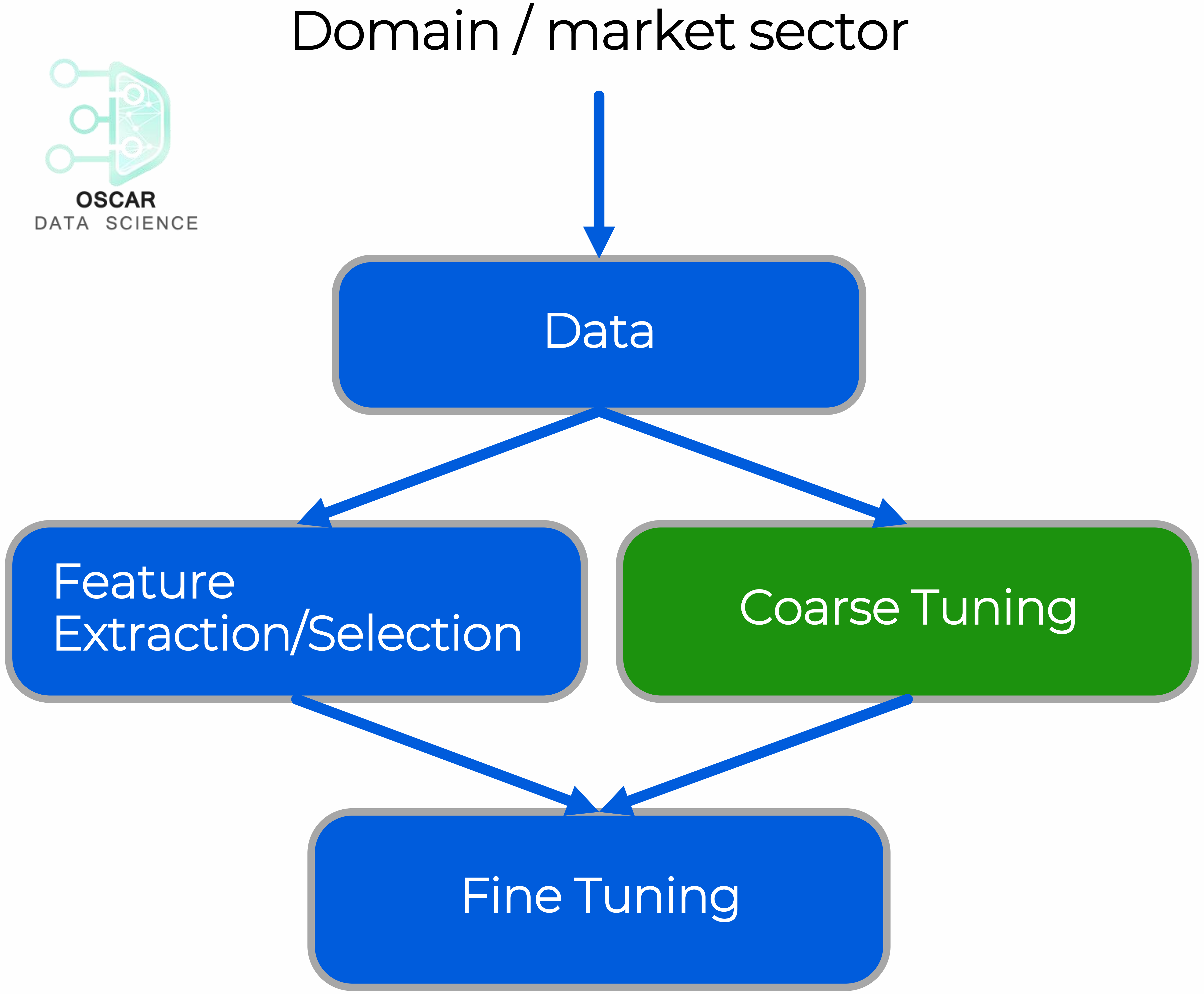
- Complex scenarios
- Many actors and variable
- Simulate and predict scenarios
- Real-time operations
- Strategic option analysis

MassiveAnalytic
Precognition Machines
www.massiveanalytic.com

OSCAR Workflow

- Coarse Tuning: deploy possibilistic algorithms not available in competitive platforms
- Feature Extraction/Selection: extend the performance and explainability of in-house solutions with new datasets
- Fine Tuning: Outperform competitive platforms in the achieved accuracies of AI pipelines

OSCAR
DATA SCIENCE



MassiveAnalytic
Precognition Machines
www.massiveanalytic.com

Problem Statement & Objective

- Real Data: multiple formats (numerical, categorical, mixed)
- Possibilisitc Decision Trees [Jenhani et al. 2008] ONLY handle Categorical features.
- Discretization of numeric features requires ignoring the possibilistic labels!

How to discretize numeric features without ignoring the uncertainty in the class labels?

Possibilistic Datasets and Non-Specificity based Possibilistic Decision Trees (NSPDT)

Jenhani I., Ben Amor N., Elouedi Z. :**Decision Trees as Possibilistic Classifiers**, *International Journal of Approximate Reasoning*, 48(3):784–807, 2008.

Standard Dataset Vs. Possibilistic Dataset

n instances; m features; c class labels

Standard dataset

	Feature 1	Feature 2	…	Feature m	Class label
1	23.5	0	…	Low	C_2
2	13.75	1	…	Average	C1
…	…	…	…	…	…
n	20	1	…	High	C_1

Ground truth: each instance belongs to only one category and we know it with certainty

Possibilistic dataset

	Feature 1	Feature 2	…	Feature m	C_1	C_2	…	C_c
1	23.5	0	…	Low	0	1	…	.4
2	13.75	1	…	Average	1	0	…	0
…	…	…	…	…	…	…	…	…
n	20	1	…	High	1	1	…	0

Ground truth: each instance belongs to only one category but we are uncertain about it

Possibilistic Dataset

Notations π

- Ω : the universe of discourse (e.g. $\{C_1, \ldots, C_c\}$)
- x : a variable with an unknown value (e.g. The class label of an object)
- ω : an element of Ω (e.g. C_2)
- L : the possibilistic scale (e.g. [0, 1])

	Feature 1	Feature 2	…	Feature m	C_1	C_2	…	C_c
1	23.5	0	…	Low	0	1	…	.4
2	13.75	1	…	Average	1	0	…	0
…	…	…	…	…	…	…	…	…
n	20	1	…	High	1	1	…	0

Possibility distribution π

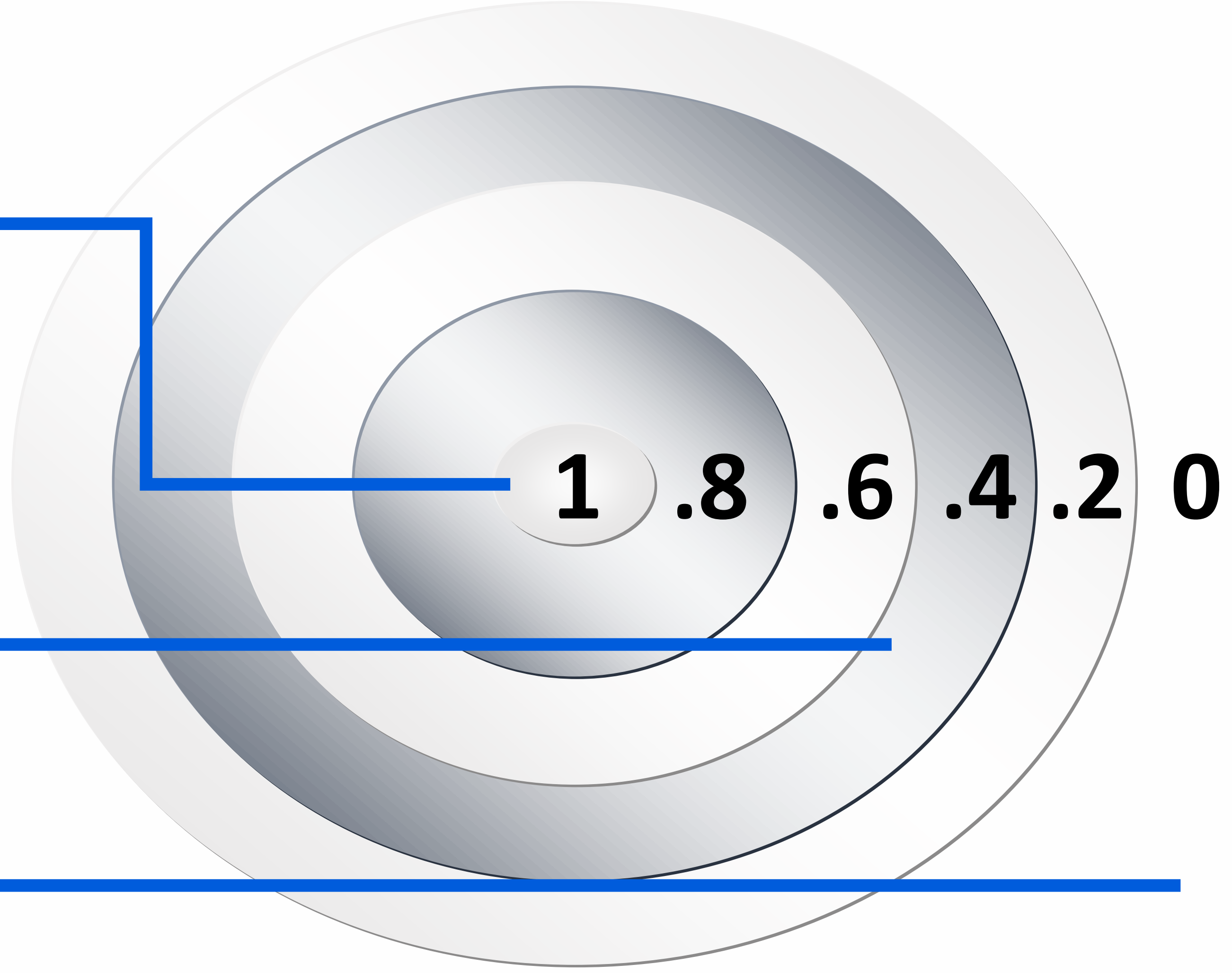
π
$: \Omega \rightarrow [0, 1]$
$\omega \rightarrow \pi_x(\omega)$

The possibility degree that x is ω

Fully possible ($\pi_C(C_2)=1$)

Possible

Totally Impossible ($\pi_C(C_1)=0$)



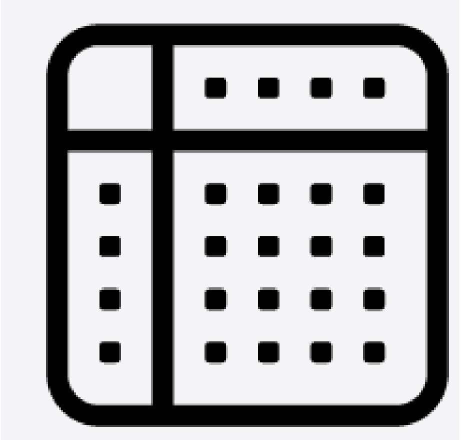
Possibilistic dataset – how to get it?

Option 1: a SME will annotate the dataset with the possibility distributions

[<u>Challenge</u>: SME should understand that a possibility distribution is not a probability distribution…]

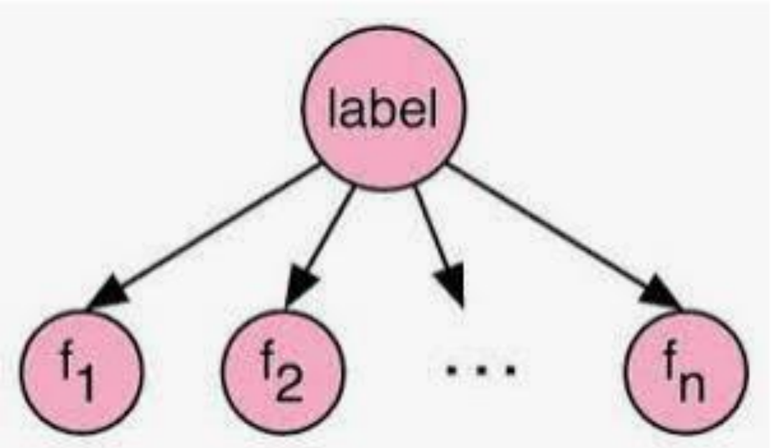
Option 2: Generate the possibilistic dataset from a standard dataset using the following procedure:

	Feature 1	Feature 2	…	Feature m	C_1	C_2	…	C_c
1	23.5	0	…	Low	0	1	…	.4
2	13.75	1	…	Average	1	0	…	0
…	…	…	…	…	…	…	…	…
n	20	1	…	High	1	1	…	0



Standard dataset

1

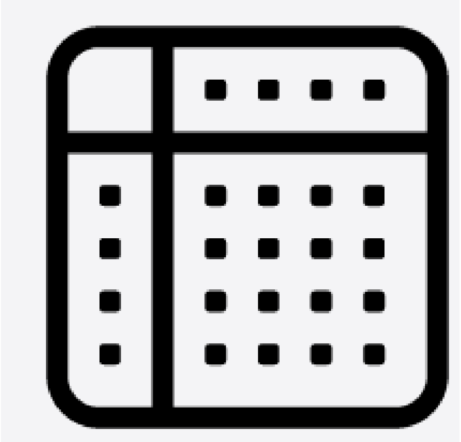


Train a Naïve Bayes model

2

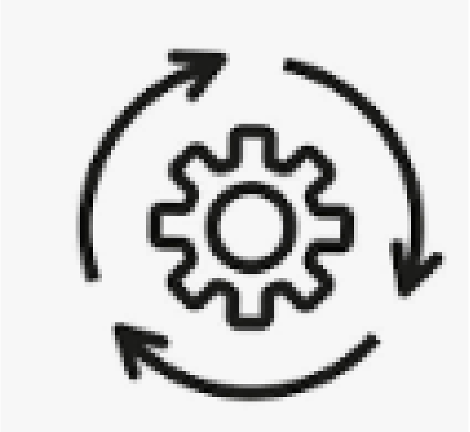
Drop the class labels from the initial standard dataset and feed it to the trained Naïve Bayes model

3



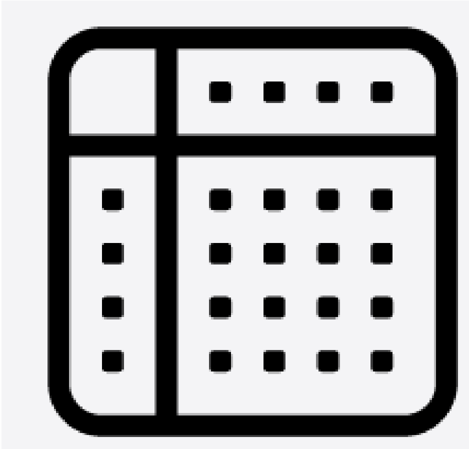
Probabilistic dataset : each label is now a probability distribution

4



Probability - Possibility transformation

5



Possibilistic dataset

- <u>Pros</u>: detects hidden class uncertainty in datasets
- <u>Cons</u>: little information compromise

Other options also exist…

NS-PDT

Building decision trees from training sets with imprecise class labels using the concept of non-specificity

Imprecise Tr

Income	Property	UnCredit	C_1	C_2	C_3
High	Greater	No	1	0.1	0.3
High	Greater	Yes	0.8	1	0.6
High	Greater	No	1	0.5	0.3
High	Less	Yes	0	1	0
Average	Greater	No	1	0	0.4
Average	Greater	Yes	0.7	1	0.2
Average	Less	No	0.7	1	0.7
Average	Less	Yes	0	1	0.3
Low	Less	No	0.5	0.5	1
Low	Less	Yes	0	0.3	1

NS-PDT: attribute selection measure

Using Non-Specificity

Partition 1

[1 0.3 0.6] [0 1 0.7]

[1 0 0] [1 0.3 1]

[0 0.6 1] [0.5 1 0.1]

→ π_{avg}^{1} [0.58 0.53 0.56]

→ π_{rep}^{1} [1 0.91 0.96]

→ $U(\pi_{rep}^{1}) = 1.49$

Partition 2

[1 1 0.3] [0.7 0 1]

[0.2 0.8 1] [1 0.3 1]

[1 0.6 0] [0 1 0]

→ π_{avg}^{2} [0.65 0.61 0.55]

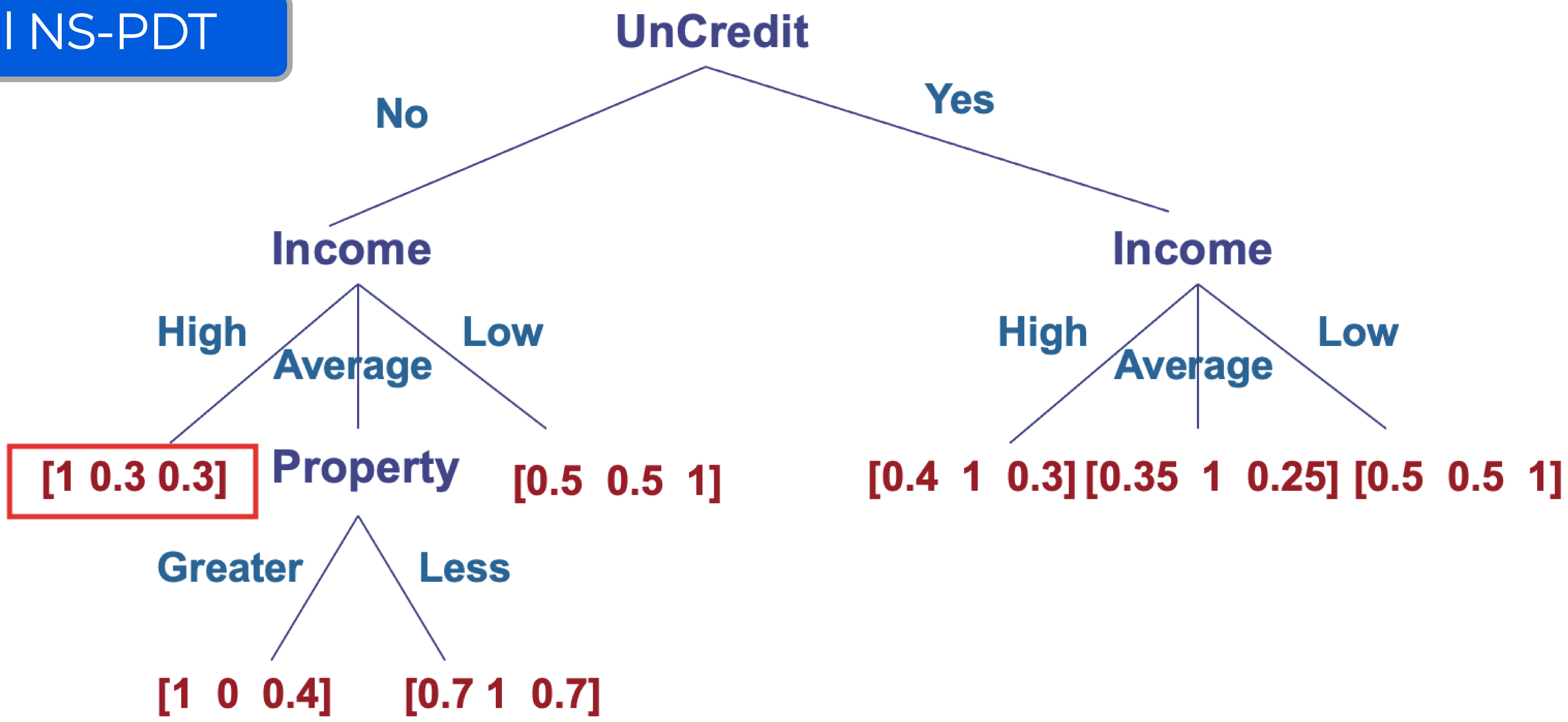
→ π_{rep}^{2} [1 0.93 0.84]

→ $U(\pi_{rep}^{2}) = 1.42$

→ Partition n°2 is more specific

Classification using NS-PDT

→ Final NS-PDT



We can predict the class of the instance: <High, Greater, No>:?

The Non-Specificity based Discretization

Standard Discretization Algorithm [Fayyad and Irani, 93]

	F_i	L
	0.1	O
	0.1	O
	0.1	O
t_1	0.3	X
	0.3	X
	0.4	X
	0.5	X
	0.5	X
t_2	0.5	O
	0.6	O
	0.6	O
	0.7	O
	0.7	O
	0.8	O
t_3	0.9	X
t_4	0.9	O
	0.9	O
t_5	0.9	X
	0.9	X

T_1

T_2

Total: 19 instance, 8: X 11: O

1. Sort instances in ASC order (by F_i)
2. Find cut point *t* that maximizes Gain(T,t)

   t: the average between 2 F_i values around the boundary point (when class label changes).

T: training partition, t: threshold that will split T into 2 sub-partitions:

- T_1 : Instances where $F_i \leq t$
- T_2 : Instances where $F_i > t$

3. Repeat from 2 for each generated sub-partition until stopping criterion is met

Possibilistic Discretization Algorithm

Cut	F_i	$\pi(X)$	$\pi(O)$	Partition
	0.1	1	0.8	T_1
	0.1	0.5	1	
t_1	0.1	0.8	1	
	0.3	1	0.2	T_2
	0.3	1	0.6	
t_2	0.4	1	0	
t_3	0.5	0.1	1	
	0.5	1	0.5	
t_4	0.5	1	0.3	
	0.6	0	1	
t_5	0.6	0.2	1	
	0.7	1	0	
	0.7	1	0.6	
	0.8	0.8	1	
t_6	0.9	1	0.7	
t_7	0.9	0.3	1	
	0.9	1	0.6	
	0.9	1	0.2	
	0.9	1	0.1	

Total: 19 instance

1. Sort instances in ASC order (by F_i)
2. Find cut point *t* that maximizes NSGain(T,t)

   Cut points: when *InfoAff* $(\pi_i, \pi_{i+1}) \leq \eta$ (hyper-parameter, default: 0.7)

   T: training partition

   t: cut point that will split T into 2 sub-partitions:

   - T_1 : Instances where $F_i \leq t$
   - T_2 : Instances where $F_i > t$

3. Repeat from 2 for each generated sub-partition until stopping criterion is met: NSGain ≤0

$$InfoAff([0.8,1],[1,0.2]) = 1 - \frac{\left(\frac{0.2+0.8}{2}\right) + 1 - (\max(\min(0.8,1), \min(1,0.2)))}{2}$$

$$= 1 - \frac{0.5+0.2}{2} = 0.65 \ (\leq 0.7 \text{ so cut point})$$

For t_1: $NSGain(T, t1 = 0.2) = U(\pi_{rep(T)}) - \left(freq_{T_1} * U(\pi_{rep(T_1)}) + freq_{T_2} * U(\pi_{rep(T2)})\right)$

$\pi_{rep}(T)$: max – normalized average possibility distribution representing T

Possibilistic Similarity Measure

Information Affinity: *[Jenhani, Benferhat and Elouedi; ECSQARU'07, IPMU'08, Foundation of Reasoning under Uncertainty, 2010]*

Definition let π_1 and π_2 be two possibility distributions on the same universe of discourse Ω. We define a measure *InfoAff* (π_1, π_2) as follows:

$$InfoAff(\pi_1,\pi_2) = 1 - \frac{d(\pi_1,\pi_2) + Inc(\pi_1 \wedge \pi_2)}{2}$$

Where $d(\pi_1,\pi_2) = \frac{1}{n}\sum_{i=1}^{n}|\pi_1(\omega_i) - \pi_2(\omega_i)|$ represents the Manhattan distance between π_1 and π_2 and $Inc(\pi_1 \wedge \pi_2)$ tells us about the degree of conflict between the two distributions.

$$Inc(\pi) = 1 - \max_{\omega\in\Omega}\{\pi(\omega)\}$$

Non-specificity Measure

U-uncertainty: ***[Klir and Folger, 1988]***

$$U\colon \mathcal{R} \to \mathbb{R}^{+}$$

where $\mathcal{R}$ denotes the set of all finite, ordered, and normal possibility distributions.

Given a possibility distribution

$$\mathrm{r} = (\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_\mathrm{n}) \text{ such that } 1 = \mathrm{r}_1 \geq \mathrm{r}_2 \geq \cdots \geq \mathrm{r}_\mathrm{n}$$

The U-uncertainty of r, U(r), can be expressed by :

$$\mathrm{U}(r) = \sum_{i=2}^{n} (r_i - r_{i+1})\, log_2\ \mathrm{i} = \sum_{i=2}^{n} r_i log_2 \frac{i}{i-1}$$

where $\mathrm{r}_{n+1} = 0$ by convention.

Algorithm: Illustration

F_i	π(X)	π(O)
0.1	1	0.8
0.1	0.5	1
0.1	0.8	1
0.3	1	0.2
0.3	1	0.6
0.4	1	0
0.5	0.1	1
0.5	1	0.5
0.5	1	0.3
0.6	0	1
0.6	0.2	1
0.7	1	0
0.7	1	0.6
0.8	0.8	1
0.9	1	0.7
0.9	0.3	1
0.9	1	0.6
0.9	1	0.2
0.9	1	0.1

T_1, T_2, t_1, t_2, t_3, t_4, t_5, t_6, t_7

19 Total: 19 instance

π_T(X)=(1+0.5+0.8+1+…+1)/19=0.774

π_T(O)=(0.8+1+1+0.2+…+0.1)/19=0.61

$\pi_{rep(T)}$(X)	$\pi_{rep(T)}$(O)
1	0.79

U(π_{rep}(T))=(1-0.79)*$\log_2$(1)+(0.79-0) *$\log_2$(2)=0.79

π_{T2}(X)=(1+1+1+0.1+…+1)/16=0.775

π_{T2}(O)=(1+0.6+0+1+…+0.1)/16=0.55

$\pi_{rep(T2)}$(X)	$\pi_{rep(T2)}$(O)
1	0.7

U(π_{rep}(T_2))=(1-0.7)*$\log_2$(1)+(0.7-0) *$\log_2$(2)=0.7

π_{T1}(X)=(1+0.5+0.8)/3=0.766

π_{T1}(O)=(0.8+1+1)/3=0.933

$\pi_{rep(T1)}$(X)	$\pi_{rep(T1)}$(O)
0.821	1

U(π_{rep}(T1))=(1-0.821)*$\log_2$(1)+(0.821-0) *$\log_2$(2)=0.821

NSGain(T, t_1=0.2)=0.79-[(16/19)*0.7+(3/19)*0.821]=0.07

Similarly, we compute NSGain(T, t_2=0.45), NSGain(T, t_3=0.5), … then select max

The consecutive selected thresholds will form the cutoff points.

Experimental Setup and Results

Datasets [UCI ML Repository]

Dataset	#instances	#classes	#features	#continuous features
Letter recognition	20 000	26	16	16
Dry Bean	13 611	7	16	16
Magic Gamma Telescope	19 020	2	10	10
Occupancy Detection	9 752	2	5	5
Spambase	4 601	2	57	57
Adult	48 842	2	14	6
Bank Marketing	45 211	2	20	10

Possibilistic versions of these datasets have been generated using Option 2 (NB + Proba-Poss transformation)

MassiveAnalytic
Precognition Machines
www.massiveanalytic.com

Results – Crisp datasets

Dataset	Classifier	Mean Acc. (Non dic.)	Mean Acc. (Std. disc.)	Mean Acc. (NS-disc.)
Letter Recognition	NB	64.01	**73.85**	70.6
	J48	**87.92**	78.75	80.59
	RF	**96.41**	93.41	92.61
	1-NN	**95.96**	91.86	89
	LogReg	91.24	**92.84**	90.14
Dry Bean	NB	89.7	**89.94**	88.5
	J48	**91.03**	90.05	90.25
	RF	**92.5**	91.47	91.14
	1-NN	**90.2**	89.25	90.2
	LogReg	**92.6**	92.3	**92.6**
Magic Gamma Telescope	NB	72.68	**78.25**	75
	J48	**85.05**	84.45	77.57
	RF	**88**	84.04	77.46
	1-NN	80.93	**81.97**	77.42
	LogReg	79.11	**84.68**	77.5
Occupancy Detection	NB	95.34	99.1	**99.33**
	J48	95.86	**99.37**	**99.37**
	RF	**99.4**	**99.4**	99.37
	1-NN	94.99	**99.37**	**99.37**
	LogReg	99.24	**99.35**	**99.35**
Spambase	NB	85.24	**90.19**	85.78
	J48	82.59	**92.82**	91.58
	RF	**95.5**	94.59	93.87
	1-NN	85.28	**93.1**	92.24
	LogReg	92.41	**94.41**	91.58
Adult	NB	83.25	**83.87**	82.35
	J48	86.1	**86.67**	85.45
	RF	85.17	**85.4**	84.75
	1-NN	79.52	83.04	**84.13**
	LogReg	85.09	**87.23**	85.58
Bank Marketing	NB	88	**88.88**	**88.88**
	J48	90.31	**90.32**	88.88
	RF	**90.38**	89.92	88.87
	1-NN	86.96	88.83	**88.87**
	LogReg	90.15	**90.4**	88.87

Crisp dataset

F1	F 2	…	F m	C_1	C_2	…	C_c
23.5	0	…	Low	0	1	0	0
13.75	1	…	Average	1	0	0	0
…	…	…	…	…	…	…	…
20	1	…	High	0	1	0	0

MassiveAnalytic
Precognition Machines
www.massiveanalytic.com

Results - Possibilistic datasets

Dataset (Discretized)	NS-PDT
Letter recognition	• CMPcc: 60.2% • InfoAffC: 0.871
Dry Bean	• CMPcc: 88.47% • InfoAffC: 0.963
Magic Gamma Telescope	• CMPcc: 81.65% • InfoAffC: 0.836
Occupancy Detection	• CMPcc: 95.9% • InfoAffC: 0.956
Spambase	• CMPcc: 97.9% • InfoAffC: 0.84
Adult	• CMPcc: 96.66% • InfoAffC: 0.96
Bank Marketing	• CMPcc: 86.82% • InfoAffC: 0.882

CMPcc: Cautious Most Plausible-based correct classification

$$CMPcc = \frac{\text{number of correctly classified instances}}{\text{total number of testing instances}} \times 100$$

InfoAffC: Information Affinity Criterion

$$\text{InfoAffC} = \frac{1}{n} \sum_{i=1}^{n} InfoAff\left(\pi_i^{\text{init}}, \pi_i^{\text{pred}}\right)$$

Conclusion and Future work

Conclusion and Future work

- A supervised discretization approach has been proposed for possibilistic (labelled) data
- Non-specificity and Information Affinity measures are used by the main building blocks of the proposed algorithm.
- The proposed pre-processing discretization approach will make it possible to use several possibilistic classifiers which were initially designed to handle categorical data only.
- With crisp data, NS discretization showed competitive results.
- More suitable for possibilistic data

Future work:

- Consider feature dependency during discretization instead of discretizing each feature individually.
- Handle unbalanced feature intervals.

Thank you!

ijenhani@pmu.edu.sa

https://sa.linkedin.com/in/ilyes-jenhani-87176124