

# Data Mining



## RULE MINING



Prof. Dr. Hikmat Ullah Khan  
Department of Information Technology

UNIVERSITY OF SARGODHA, SARGODHA

# Lesson from Holy Quran

2

## Quran Verses About Forgiveness

أَفَلَا يَتُوبُونَ إِلَى اللَّهِ وَيَسْتَغْفِرُونَهُ  
وَاللَّهُ غَفُورٌ رَحِيمٌ

Will they not turn to Allah in  
**repentance** and seek His **forgiveness**?  
And **Allah is All-Forgiving, Most  
Merciful**

—Quran 5:74

# Rule-Based Classifier

3

- Classify records by using a collection of “if...then...” rules
- Rule:  $(Condition) \rightarrow y$ 
  - ▣ where
    - *Condition* is a conjunctions of attributes
    - *y* is the class label
  - ▣ *LHS*: rule antecedent or condition
  - ▣ *RHS*: rule consequent
  - ▣ Examples of classification rules:
    - $(\text{Blood Type}=\text{Warm}) \wedge (\text{Lay Eggs}=\text{Yes}) \rightarrow \text{Birds}$
    - $(\text{Taxable Income} < 50\text{K}) \wedge (\text{Refund}=\text{Yes}) \rightarrow \text{Cheat}=\text{No}$

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
human	warm	yes	no	no	mammals
python	cold	no	no	no	reptiles
salmon	cold	no	no	yes	fishes
whale	warm	yes	no	yes	mammals
frog	cold	no	no	sometimes	amphibians
komodo	cold	no	no	no	reptiles
bat	warm	yes	yes	no	mammals
pigeon	warm	no	yes	no	birds
cat	warm	yes	no	no	mammals
leopard shark	cold	yes	no	yes	fishes
turtle	cold	no	no	sometimes	reptiles
penguin	warm	no	no	sometimes	birds
porcupine	warm	yes	no	no	mammals
eel	cold	no	no	yes	fishes
salamander	cold	no	no	sometimes	amphibians
gila monster	cold	no	no	no	reptiles
platypus	warm	no	no	no	mammals
owl	warm	no	yes	no	birds
dolphin	warm	yes	no	yes	mammals
<sup>4</sup> eagle	warm	no	yes	no	birds

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
human	warm	yes	no	no	mammals
python	cold	no	no	no	reptiles
salmon	cold	no	no	yes	fishes
whale	warm	yes	no	yes	mammals
frog	cold	no	no	sometimes	amphibians
komodo	cold	no	no	no	reptiles
bat	warm	yes	yes	no	mammals
pigeon	warm	no	yes	no	birds
cat	warm	yes	no	no	mammals
leopard shark	cold	yes	no	yes	fishes
turtle	cold	no	no	sometimes	reptiles
penguin	warm	no	no	sometimes	birds
porcupine	warm	yes	no	no	mammals
eel	cold	no	no	yes	fishes
salamander	cold	no	no	sometimes	amphibians
gila monster	cold	no	no	no	reptiles
platypus	warm	no	no	no	mammals
owl	warm	no	yes	no	birds
dolphin	warm	yes	no	yes	mammals
eagle	warm	no	yes	no	birds

**R1: (Give Birth = no)  $\wedge$  (Can Fly = yes)  $\rightarrow$  Birds**

**R2: (Give Birth = no)  $\wedge$  (Live in Water = yes)  $\rightarrow$  Fishes**

**R3: (Give Birth = yes)  $\wedge$  (Blood Type = warm)  $\rightarrow$  Mammals**

**R4: (Give Birth = no)  $\wedge$  (Can Fly = no)  $\rightarrow$  Reptiles**

**R5: (Live in Water = sometimes)  $\rightarrow$  Amphibians**

# Application of Rule-Based Classifier

6

- A rule  $r$  **covers** an instance  $x$  if the attributes of the instance satisfy the condition of the rule

**R1: (Give Birth = no)  $\wedge$  (Can Fly = yes)  $\rightarrow$  Birds**

**R2: (Give Birth = no)  $\wedge$  (Live in Water = yes)  $\rightarrow$  Fishes**

**R3: (Give Birth = yes)  $\wedge$  (Blood Type = warm)  $\rightarrow$  Mammals**

**R4: (Give Birth = no)  $\wedge$  (Can Fly = no)  $\rightarrow$  Reptiles**

**R5: (Live in Water = sometimes)  $\rightarrow$  Amphibians**

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
hawk	warm	no	yes	no	?
grizzly bear	warm	yes	no	no	?

**The rule R1 covers a hawk  $\Rightarrow$  Bird**

**The rule R3 covers the grizzly bear  $\Rightarrow$  Mammal**

# Rule Coverage and Accuracy

7

- Coverage of a rule:
  - ▣ Fraction of records that satisfy the antecedent of a rule
- Accuracy of a rule:
  - ▣ Fraction of records that satisfy both the antecedent and consequent of a rule

<i>Tid</i>	Refund	Marital Status	Taxable Income	Class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

(Status=Single) → No

Coverage = 40%, Accuracy = 50%

# How does Rule-based Classifier Work?

8

- R1: (Give Birth = no)  $\wedge$  (Can Fly = yes)  $\rightarrow$  Birds
- R2: (Give Birth = no)  $\wedge$  (Live in Water = yes)  $\rightarrow$  Fishes
- R3: (Give Birth = yes)  $\wedge$  (Blood Type = warm)  $\rightarrow$  Mammals
- R4: (Give Birth = no)  $\wedge$  (Can Fly = no)  $\rightarrow$  Reptiles
- R5: (Live in Water = sometimes)  $\rightarrow$  Amphibians

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
lemur	warm	yes	no	no	?
turtle	cold	no	no	sometimes	?
dogfish shark	cold	yes	no	yes	?

- A lemur triggers rule R3, so it is classified as a mammal
- A turtle triggers both R4 and R5
- A dogfish shark triggers none of the rules



# Ordered Rule Set

9

- Rules are rank ordered according to their priority
  - ▣ An ordered rule set is known as a decision list
  - ▣ What is priority?
- When a test record is presented to the classifier
  - ▣ It is assigned to the class label of the highest ranked rule it has triggered
  - ▣ If none of the rules fired, it is assigned to the default class

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
turtle	cold	no	no	sometimes	?

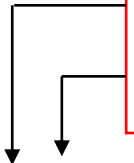
R1: (Give Birth = no)  $\wedge$  (Can Fly = yes)  $\rightarrow$  Birds

R2: (Give Birth = no)  $\wedge$  (Live in Water = yes)  $\rightarrow$  Fishes

R3: (Give Birth = yes)  $\wedge$  (Blood Type = warm)  $\rightarrow$  Mammals

R4: (Give Birth = no)  $\wedge$  (Can Fly = no)  $\rightarrow$  Reptiles

R5: (Live in Water = sometimes)  $\rightarrow$  Amphibians



# Rule Ordering Schemes

10

- Rule-based ordering
  - ▣ Individual rules are ranked based on quality
- Class-based ordering
  - ▣ Rules that belong to the same class appear together

## Rule-based Ordering

(Refund=Yes) ==> No

(Refund=No, Marital Status={Single,Divorced},  
Taxable Income<80K) ==> No

(Refund=No, Marital Status={Single,Divorced},  
Taxable Income>80K) ==> Yes

(Refund=No, Marital Status={Married}) ==> No

## Class-based Ordering

(Refund=Yes) ==> No

(Refund=No, Marital Status={Single,Divorced},  
Taxable Income<80K) ==> No

(Refund=No, Marital Status={Married}) ==> No

(Refund=No, Marital Status={Single,Divorced},  
Taxable Income>80K) ==> Yes

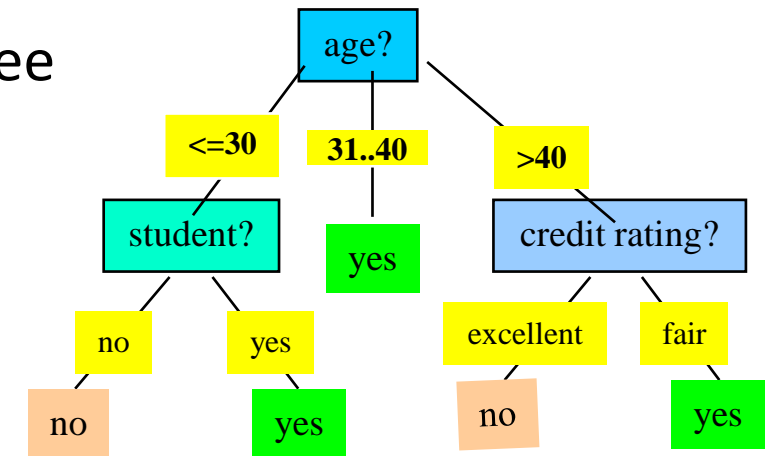
# Rule Extraction from a Decision Tree

11

- Rules are *easier to understand* than large trees
- One rule is created *for each path* from the root to a leaf
- Each attribute-value pair along a path forms a conjunction: the leaf holds the class prediction

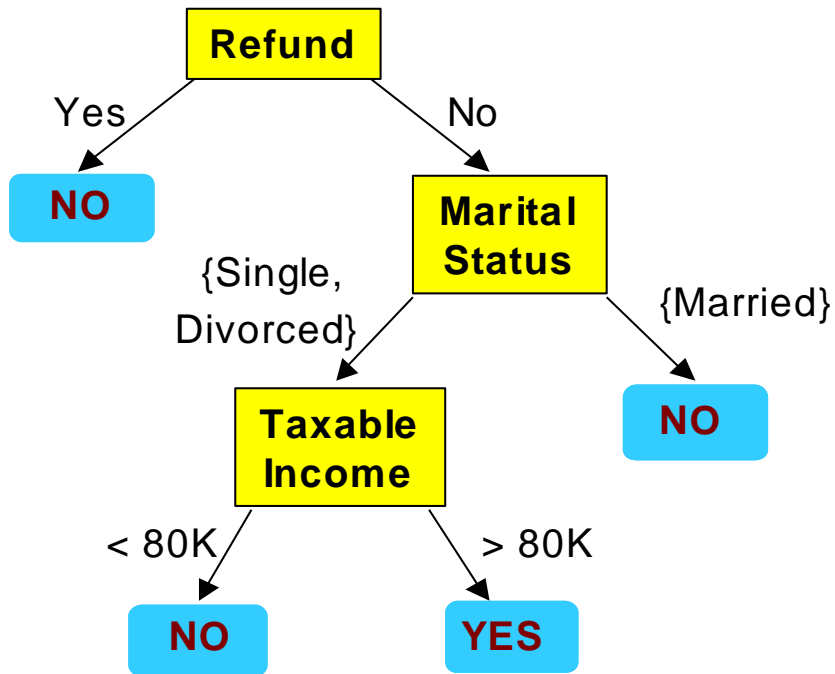
Example: Rule extraction from our *buys\_computer* decision-tree

IF <i>age</i> = young AND <i>student</i> = no	THEN <i>buys_computer</i> = no
IF <i>age</i> = young AND <i>student</i> = yes	THEN <i>buys_computer</i> = yes
IF <i>age</i> = mid-age	THEN <i>buys_computer</i> = yes
IF <i>age</i> = old AND <i>credit_rating</i> = excellent	THEN <i>buys_computer</i> = no
IF <i>age</i> = old AND <i>credit_rating</i> = fair	THEN <i>buys_computer</i> = yes



# From Decision Trees To Rules

12



## Classification Rules

(Refund=Yes) ==> No

(Refund=No, Marital Status={Single, Divorced}, Taxable Income<80K) ==> No

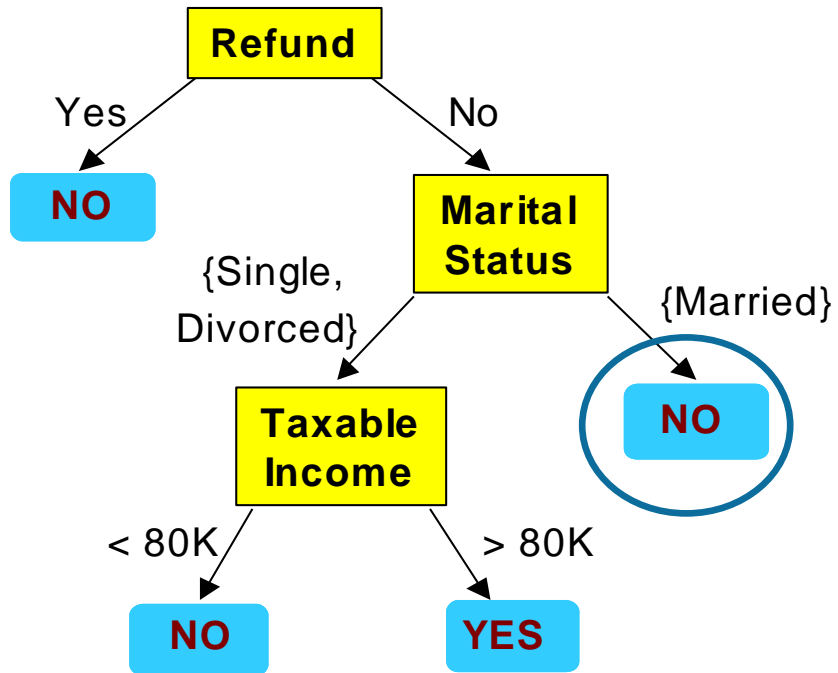
(Refund=No, Marital Status={Single, Divorced}, Taxable Income>80K) ==> Yes

(Refund=No, Marital Status={Married}) ==> No

Rule set contains as much information as the tree

# Rules Can Be Simplified

13



Initial Rule:  $(\text{Refund}=\text{No}) \wedge (\text{Status}=\text{Married}) \rightarrow \text{No}$

Simplified Rule:  $(\text{Status}=\text{Married}) \rightarrow \text{No}$

<i>Tid</i>	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

# Advantages

14

- Rules are Easy to generate, understandable
- Rules are highly expressive/interpretable.
- Predicts class labels very quickly in significantly less time
- Results are comparable to that of a decision tree.

# Rule Extraction Algorithms- Task (optional but recommended)

15

- There are **some algorithms** that are used to Extract Rules from Data
- Try to check Tools for such algorithms
  - ▣ WEKA
  - ▣ RapidMiner
- Try to use languages like
  - ▣ R
  - ▣ Python
- Try to Use Some Datasets you like
  - ▣ Coronavirus
  - ▣ Any other dataset you like



The capacity to learn is  
a *gift*; the ability to  
learn is a *skill*; the  
willingness to learn is a  
*choice*.

Brian Herbert