Trendwise Analytics

GOOD SOLUTIONS
FOR YOUR BUSINESS!

Introduction to decision trees

S.Mohan Kumar

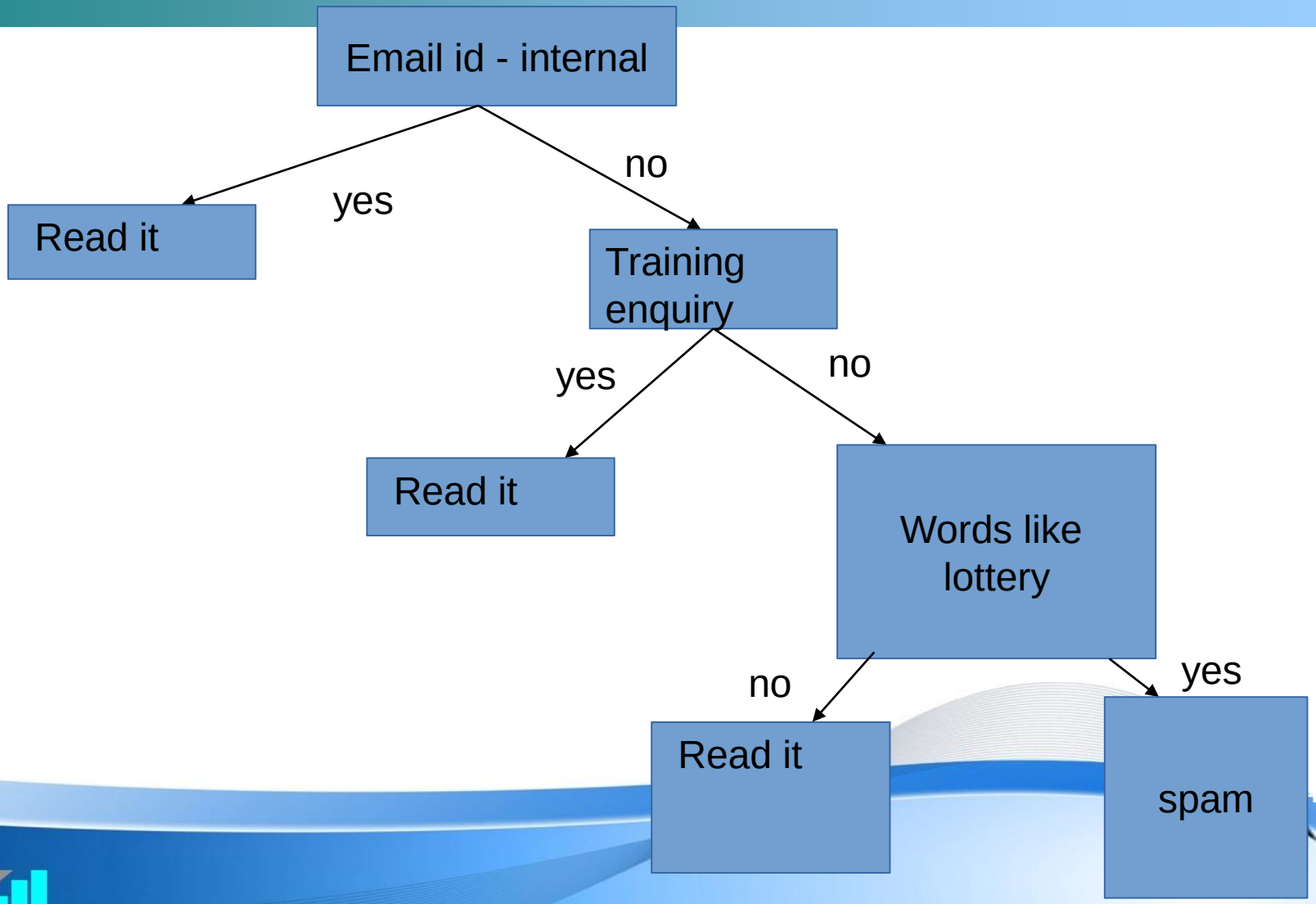
Contents

- Introduction to decision trees
- Demo and Hands on
- Q&A

Introduction

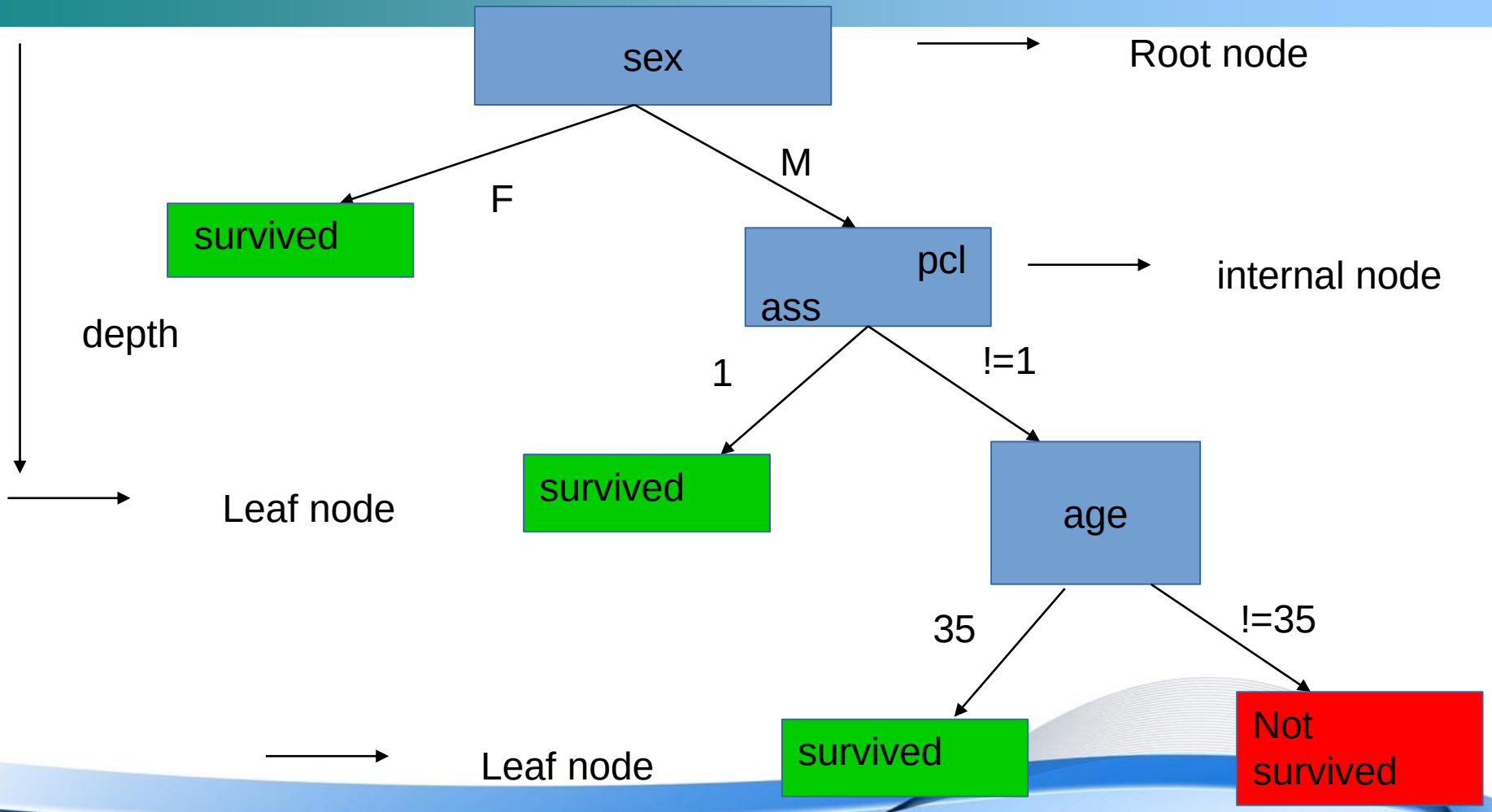
- A decision tree creates a type of a flow chart which consists of nodes (referred to as leaves) and a set of decisions to be made based off of nodes (also known as branches)
- This is a classification algorithm
- They are very simple to understand and explain
- They work with small amount of data as well
- This is used in business for credit risk modelling etc.
- There are other business applications like in pharmacy , manufacturing etc.
-

Simple example



Survived	Pclass	Sex	Age	SibSp	Parch
0	3	male	22	1	0
1	1	female	38	1	0
1	3	female	26	0	0
1	1	female	35	1	0
1	3	male	35	0	0
0	3	male		0	0
1	1	male	54	0	0

Decision tree



Advantages of decision trees

- Algorithm is easy to understand and explain
- Can be used for classification as well as regression
- Fast and efficient
- Rule based - hence easy to program
- SQL queries can be written to implement it
-
-

Disadvantages of decision trees

- Does not work very well with categorical data
- Calculations can be come very complicated when dealing with lots of linked outcomes
-
-
-

Random forest

- Random forest algorithm is an ensemble learning method of classification and regression
- It operates by constructing a multitude of decision trees (hence the term forest)
- In case of classification it is the mode of the classes
- In case of regression it is the mean of the outputs
- Advantages and disadvantages are same as decision trees.
-
- Demo
-