Data Mining

Classification IV - Random Forests (Part C)

Dr. Jason T.L. Wang, Professor

Department of Computer Science

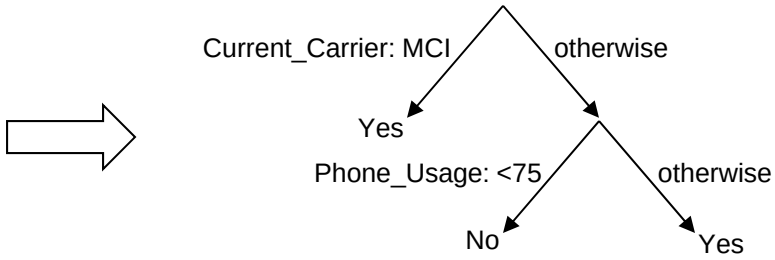
New Jersey Institute of Technology

/

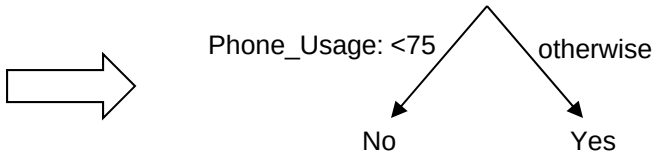
Where am I?

- Part A explains how the random forests algorithm works.
- Part B presents an example to show how the algorithm grows a tree.
- Part C shows how the algorithm grows three trees and how the algorithm uses the three trees to classify an unlabeled test record.

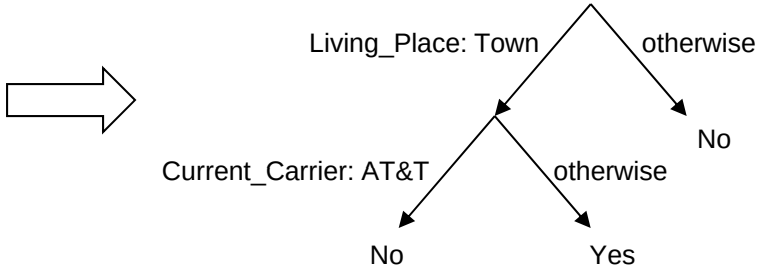
Phone_Usage	Income_Source	Living_Place	Current_Carrier	Change_Plan
>150	1	Town	AT&T	Yes
<75	1	Town	AT&T	No
<75	2	City	Sprint	No
75...150	2	City	MCI	Yes
75...150	2	City	Sprint	Yes
75...150	1	Town	MCI	Yes



Phone_Usage	Income_Source	Living_Place	Current_Carrier	Change_Plan
75…150	1	Town	MCI	Yes
75…150	2	City	AT&T	Yes
<75	1	City	Sprint	No
<75	2	Town	AT&T	No
>150	1	City	MCI	Yes



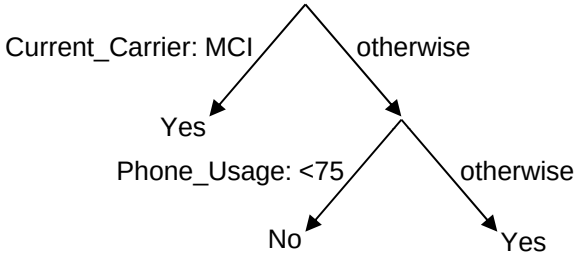
Phone_Usage	Income_Source	Living_Place	Current_Carrier	Change_Plan
>150	2	Town	Sprint	Yes
>150	2	Town	MCI	Yes
75…150	2	Town	MCI	Yes
>150	2	City	AT&T	No
>150	2	City	MCI	No
75…150	2	Town	AT&T	No
<75	2	Town	AT&T	No

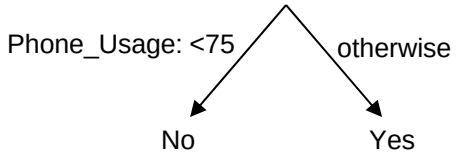


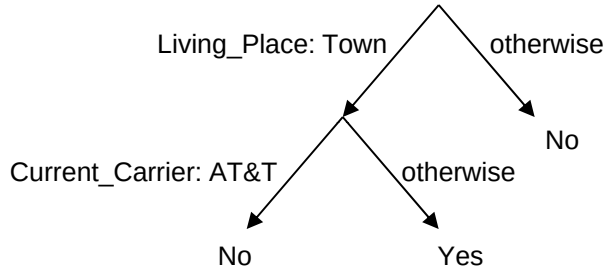
Here we show 3 randomly picked training sets and the CART tree built for each training set.

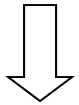
Classifying an unlabeled test record

Phone_Usage	Income_Source	Living_Place	Current_Carrier	Change_Plan
<75	2	Town	MCI	?

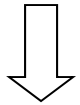




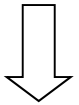




Yes



No



Yes

The majority vote is “Yes”, so we classify the unlabeled test record to the “Yes” class.

End of
Random Forests Module (Part C)