Practice Problem on Random Forest

The ideal living condition depends on the temperature, wind, and humidity of the day. The recorded temperature, wind, and humidity are given in the following dataset along with the living status. We are trying to train a smart-home agent on the following dataset using the Random Forest algorithm to display the living status of a smart home.

Construct the 1st decision tree using the instances of Day1 to Day5 and considering Temperature and Wind features; the 2nd decision tree using the instances of Day2 to Day6 and considering Wind and Humidity features; Please show all the necessary calculations to build the decision trees considering 2 as the expansion threshold.

	Day	***Temperature***	***Wind***	***Humidity***	***Living status?***
	Day1	20	Weak	Normal	Ideal
	Day2	22	Strong	Abnormal	Non-ideal
	Day3	30	Strong	Normal	Ideal
	Day4	24	Weak	Abnormal	Ideal
	Day5	32	Strong	Normal	Non-ideal
	Day6	34	Weak	Normal	Ideal
	Day7	36	Weak	Abnormal	Non-ideal

If the recorded Temperature, Wind, and Humidity of Day 8 are '33', 'Strong', and 'Abnormal' respectively. Infer the decision of the smart-home agent regarding the living status of Day 8 using the bagging procedure of the Random Forest classifier.

Day	*Temp*	*Wind*	*Humidity*	Living status?
D1	20	Weak	Normal	Ideal
D2	22	Strong	Abnormal	Non-ideal
D3	30	Strong	Normal	Ideal
D4	24	Weak	Abnormal	Ideal
D5	32	Strong	Normal	Non-ideal
D6	34	Weak	Normal	Ideal
D7	36	Weak	Abnormal	Non-ideal

1st Bootstrapped Dataset

Day	*Temp*	*Wind*	*Humidity*	Living status?
D1	20	Weak	Normal	Ideal
D2	22	Strong	Abnormal	Non-ideal
D3	30	Strong	Normal	Ideal
D4	24	Weak	Abnormal	Ideal
D5	32	Strong	Normal	Non-ideal

Feature subset:

Day	*Temp*	*Wind*		Living status?
D1	20	Weak		Ideal
D2	22	Strong		Non-ideal
D3	30	Strong		Ideal
D4	24	Weak		Ideal
D5	32	Strong		Non-ideal

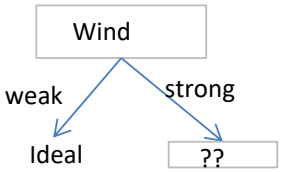
	Temp	**Class**		**Number of Left side instances with "Ideal class"**	**Number of Left side instances with "Non-Ideal"**	**GINI score of Left side instances**	**Number of Right side instances with "Ideal"**	**Number of Right side instances with "Non-ideal Class"**	**GINI score of Right side instances**	**GINI induced by the separator**
D1	20	Ideal	**Considering Temp< 21 as the separator**	1	0	0	3	3	**0.5**	0.6
D2	22	Non-ideal	**Considering Temp<23 as the separator**	1	1	0.5	3	2	**0.48**	0.68
D4	24	Ideal	**Considering Temp<27 as the separator**	2	1	0.4444	2	2	**0.5**	0.666667
D3	30	Ideal	**Considering Temp<31 as the separator**	3	1	0.375	1	2	**0.44444444**	0.566667
D5	32	Non-ideal							Minimum GINI	0.5667

GINI Impurity of Temp i.e., (Temp< 31)= 0.5666667

Living Status:

GINI Impurity of Living Status= 0.48

Wind	*Ideal*	*Non-ideal*	*GINI Impurity*
Weak(2)	2	0	0
Strong(3)	1	2	0.444444444
GINI Impurity of Wind=			0.266666667

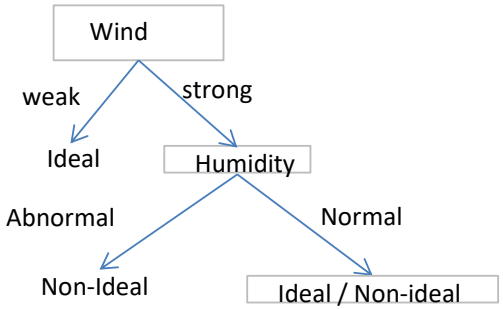


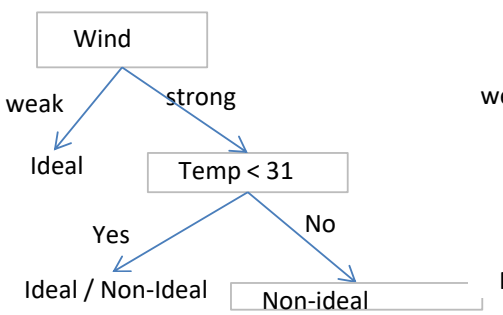
Day	*Temp*	*Wind*	*Humidity*	Living status?
D2	22	Strong	Abnormal	Non-ideal
D3	30	Strong	Normal	Ideal
D5	32	Strong	Normal	Non-ideal

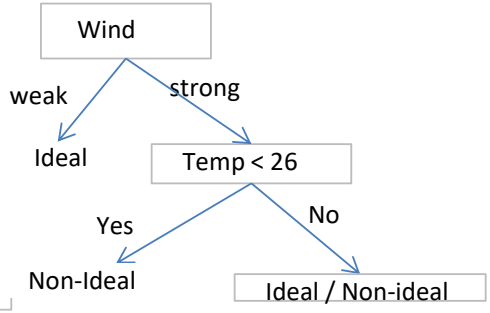
	Temp	Class		Number of Left side instances with "Ideal class"	Number of Left side instances with "Non-Ideal"	GINI score of Left side instances	Number of Right side instances with "Ideal"	Number of Right side instances with "Non-ideal Class"	GINI score of Right side instances	GINI induced by the separator
D2	22	Non-ideal	**Considering Temp<26 as the separator**	0	1	0	1	1	**0.5**	0.333333
D3	30	Ideal	**Considering Temp<31 as the separator**	1	1	0.5	0	1	**0**	0.333333
D5	32	Non-ideal								
									Minimum GINI	0.3333

GINI Impurity of Temp i.e., (Temp< 31) OR (Temp<26)= 0.3333333

Humidity	*Ideal*	*Non-ideal*		*GINI Impurity*
Normal (2)	1	1		0.5
Abnormal(1)	0	1		0
	GINI Impurity of Humidity=			0.333333333







2nd Bootstrapped Dataset

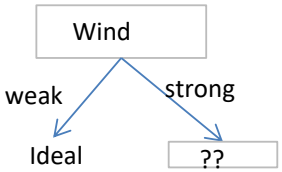
Day	***Temp***	***Wind***	***Humidity***	**Living status?**
D2	22	Strong	Abnormal	Non-ideal
D3	30	Strong	Normal	Ideal
D4	24	Weak	Abnormal	Ideal
D5	32	Strong	Normal	Non-ideal
D6	34	Weak	Normal	Ideal

Feature subset:

Day		*Wind*	*Humidity*	Living status?
D2		Strong	Abnormal	Non-ideal
D3		Strong	Normal	Ideal
D4		Weak	Abnormal	Ideal
D5		Strong	Normal	Non-ideal
D6		Weak	Normal	Ideal

Wind	*Ideal*	*Non-ideal*	*GINI Impurity*
Weak(2)	2	0	0
Strong(3)	1	2	0.444444444
	GINI Impurity of Wind=		0.266666667

Humidity	*Ideal*	*Non-ideal*	*GINI Impurity*
Normal (3)	2	1	0.444444444
Abnormal(2)	1	1	0.5
	GINI Impurity of Humidity=		0.466666667

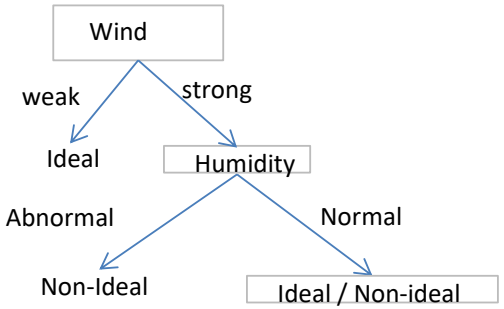


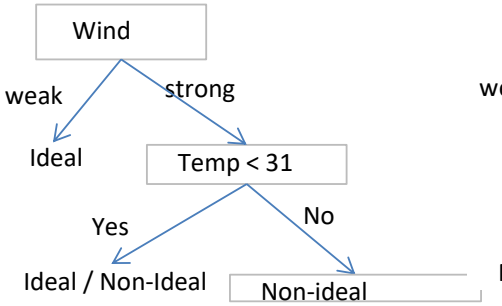
Day	*Temp*	*Wind*	*Humidity*	Living status?
D2	22	Strong	Abnormal	Non-ideal
D3	30	Strong	Normal	Ideal
D5	32	Strong	Normal	Non-ideal

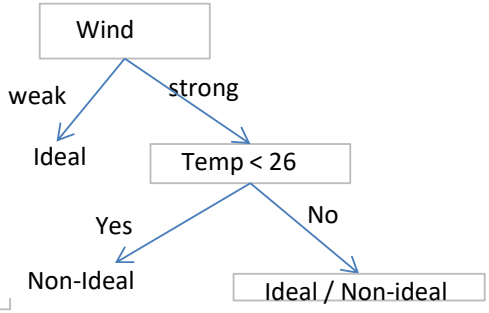
	Temp	**Class**		**Number of Left side instances with "Ideal class"**	**Number of Left side instances with "Non-Ideal"**	**GINI score of Left side instances**	**Number of Right side instances with "Ideal"**	**Number of Right side instances with "Non-ideal Class"**	**GINI score of Right side instances**	**GINI induced by the separator**
D2	22	Non-ideal	**Considering Temp<26 as the separator**	0	1	0	1	1	**0.5**	0.333333
D3	30	Ideal	**Considering Temp<31 as the separator**	1	1	0.5	0	1	**0**	0.333333
D5	32	Non-ideal								
									Minimum GINI	0.3333

GINI Impurity of Temp i.e., (Temp< 31) OR (Temp<26)= 0.3333333

Humidity	*Ideal*	*Non-ideal*	*GINI Impurity*
Normal (2)	1	1	0.5
Abnormal(1)	0	1	0
GINI Impurity of Humidity=			0.333333333







Query:

The recorded Temperature, Wind, and Humidity of Day 8 are '33', 'Strong', and 'Abnormal'

1st Tree: Non-ideal

2nd Tree: Non-ideal

3rd Tree: Ideal or Non-ideal

Answer: Non-ideal