Predicting Weather on Your Own with Decision Tree

Atika Firdaus / 13514009
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13514009@std.stei.itb.ac.id

***Abstract*—For some people, weather is one of the most important things to know before he/she starts his/her day. With information about particular day's weather, people can decide which activity they shall do and they shall not do. This paper will explain how to predict weather using decision tree.**

***Keywords*—weather, prediction, decision tree.**

I. INTRODUCTION

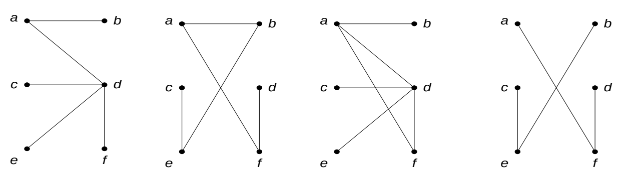
The word "Weather" refers to condition of the air on earth at a given place and time.[1] The difference between weather and climate is weather relates to the condition of the air on earth in narrow area in short term, unlike climate which relates to more bigger area in long term.

People's activities have been depending on weather, and will always be depending on it. Then, information about weather is essential, people need it. Mostly, weather information are published through media such as television and newspaper, as known as weather prediction. The information they give including the day's temperature, air humidity, rainfall intensity, and wind speed. Moreover, with nowadays' technology, people can get weather information easily from their smartphones or any other gadgets. But, it will be easier if we have the ability to predict today's weather on our own.

There are several aspects that influence the weather: air pressure, air masses, wind, temperature, humidity, precipitation (in the form of clouds), etc. But, this paper will point out more about how to predict weather by observing clouds.

II. BASIC THEORIES

Tree is defined as a connected yet undirected graph with no simple circuit.[2]



Tree Tree Not a Tree Not a Tree

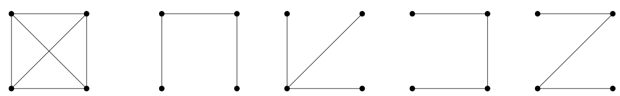
Figure 2.1

Theorem. Assume that $G = (V, E)$ is a simple undirected graph with n nodes. Then, all these statements are equivalent:

1. G is a tree.
2. Any pair of vertices in G are connected by exactly one path.
3. G is connected with $m = n - 1$ edges.
4. G has no circuits with $m = n - 1$ edges.
5. G has no circuits and an addition of an edge to G will only form one circuit.

A. Spanning Tree

Let G be a tree. A spanning tree of G is a subgraph of G that is a tree containing every node of G. Spanning tree is obtained by removing circuit in the graph.



G T1 T2 T3 T4

Figure 2.2

Every connected graph consists at least one spanning tree. Unconnected graph with k components consists k spanning forests.

B. Application of Spanning Tree

- The minimum amount of roads which connect all the cities so that each city stay connected with one another.
- Message routing in computer network.

C. Minimum Spanning Tree

- A weighted connected graph may have more than one spanning tree.
- Spanning tree with the sum of the weights of the edges of the tree is minimized is called a minimum spanning tree.

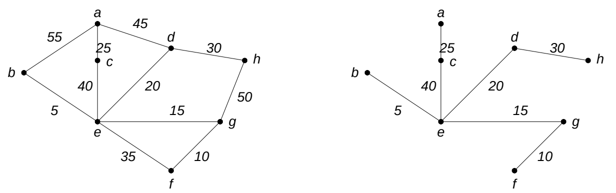


Figure 2.3

There are two algorithms for constructing minimum spanning tree, Prim's Algorithm and Kruskal's Algorithm.[3]

Prim's Algorithm

Assume *G* is a connected-undirected graph with *n* nodes,

Step 1 : Take edges of graph *G* with the minimum weight, add them to *T*

Step 2 : Take an edge (*u*, *v*) of minimum weight incident to a node in *T*, and not forming a simple circuit in *T* if added to *T*.

Step 3 : Repeat step 2 *n* – 2 times.

> **procedure** *Prim* (*G*: weighted connected undirected graph with *n* vertices)
>
> *T* := a minimum-weight edge
> **for** *i* := 1 **to** *n* − 2
>   *e* := an edge of minimum weight incident to a vertex in *T* and not forming a
>   simple circuit in *T* if added to *T*
>   *T* := *T* with *e* added
> **return** *T* {*T* is a minimum spanning tree of *G*}

Example:

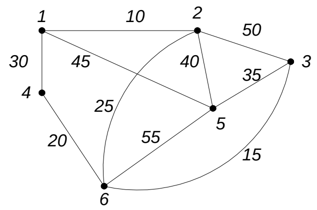


Figure 2.4

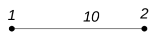
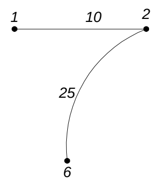
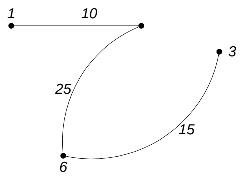
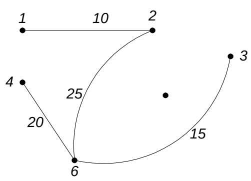
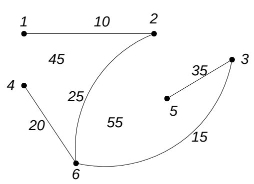
Step	Edge	Weight	Spanning Tree
1	(1, 2)	10	
2	(2, 6)	25	
3	(3, 6)	15	
4	(4, 6)	20	
5	(3, 5)	35	

Figure 2.5

Result of minimum spanning tree:

Weight = 10 + 25 + 15 + 20 + 35
= 105

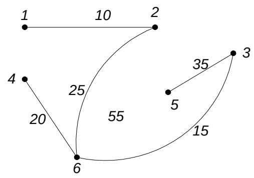


Figure 2.6

Kruskal's Algorithm

Assume *G* is a connected-undirected graph with *n* nodes,

Step 1 : Sort all the edges in *G* by their weight, in ascending order.

Step 2 : Take an edge (*u*, *v*) of minimum weight which will not form a simple circuit in *T*, add (*u*, *v*) to *T*.

Step 3 : Repeat step 2 *n* – 1 times.

> **procedure** *Kruskal* (*G*: weighted connected undirected graph with *n* vertices)
>
> *T* := empty graph
> **for** *i* := 1 **to** *n* − 1
>   *e* := any edge in *G* with smallest weight that does not form a simple circuit
>   when added to *T*
>   *T* := *T* with *e* added
> **return** *T* {*T* is a minimum spanning tree of *G*}

Example:

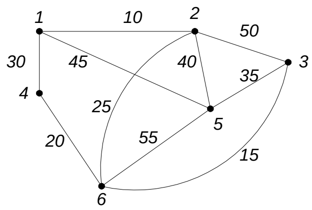
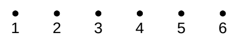
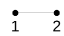
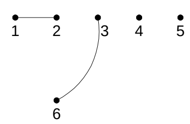
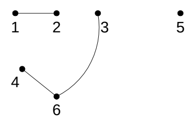


Figure 2.7

Edges in ascending order:

Edge	(1,2)	(3,6)	(4,6)	(2,6)	(1,4)
Weight	10	15	20	25	30
Edge	(3,5)	(2,5)	(1,5)	(2,3)	(5,6)
Weight	35	40	45	50	55

Step	Edge	Weight	Spanning Forest
0			
1	(1, 2)	10	
2	(3, 6)	15	
3	(4, 6)	20	

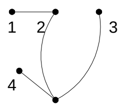
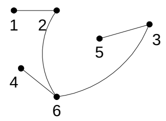
4	(2, 6)	25	
5	(1, 4)	30	ditolak
6	(3, 5)	35	

Figure 2.8

Result of minimum spanning tree:
Weight = 10 + 25 + 15 + 20 + 35
= 105

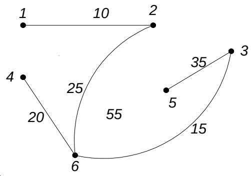


Figure 2.9

D. Rooted Tree

A rooted tree is a tree in which one node has been designated as the root and every edge is directed away from the root.

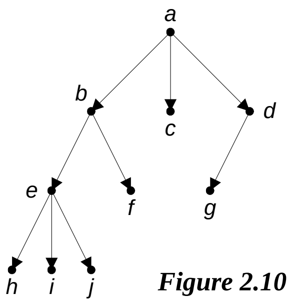
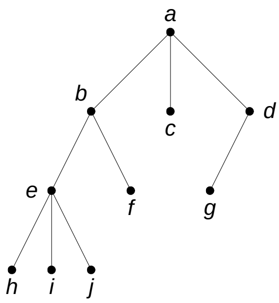


Figure 2.10



E. Terminology for Trees

The terminology for trees are originated from botanical and genealogical. Assume that *T* is a rooted tree. If *v* is a node in *T* other than root, the parent of *v* is the unique node *u*, such that there is a directed edge for *u* to *v*. When *u* is the parent of *v*, then *v* is child of *u*. Nodes with same parent are called siblings. The ancestors of a node other than root are nodes in the path from the root to this node, including the root, but excluding the node itself. The descendants of node *u* are those nodes that have *u* as an ancestor.

Degree of node *u* is the amount of child of *u*. Nodes with 0 degree are called leaf. Nodes other than root which is not leaf, is called internal nodes. Height or depth is the maximum level of the tree. Level, is explained in figure 2.11.

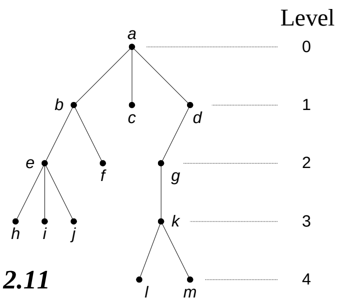


Figure 2.11

If *u* is a node in a tree, the subtree with *u* as its root is the subgraph of the tree consisting of *u* and its descendants and all edges incident to these descendants.

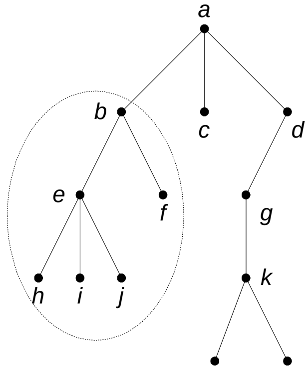
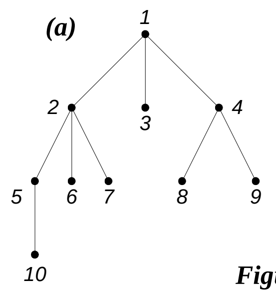


Figure 2.12

F. Ordered Tree

Ordered tree is a kind of tree in which the children of each internal nodes are linearly ordered.

(a)



(b)

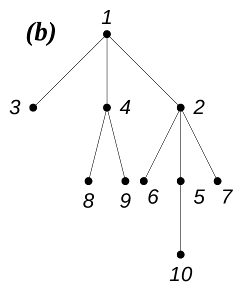


Figure 2.13

Figure 2.13 (a) and (b) are two different ordered trees.

G. *n*-ary and Binary Tree

n-ary tree is a tree in which the maximum degree of each nodes is *n*. *n*-ary tree is considered full when every node's degree is exactly *n*.

Binary tree is a *n*-ary tree with *n* = 2. Every node in binary tree has no more than 2 children. If a node has 2 childern, the first child called left child, and the second one called right child. Because of the children are differed, then binary tree is also an ordered tree.

A balanced-tree is a tree which the height difference between right subtree and left subtree at each node is not greater than 1.

H. Decision Tree as an Application of Binary Tree

Figure 2.14

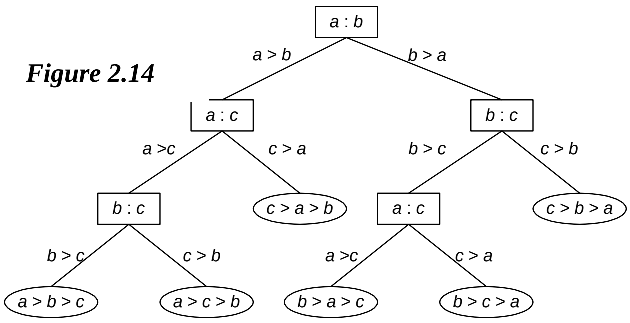


Figure 2.1.14 is decision tree for sorting three elements.

III. CLASSIFYING CLOUDS WITH DECISION TREE

Clouds are classified according to its altitude and its appearance.[4]

By its altitude, clouds are classified into two:

1. Cirro

   Cirro in Latin means curl of hair.[5] Cirro clouds are clouds in the high altitude range, above 6 kilometers. [6]
2. Alto

   Alto clouds are clouds in the middle altitude range, between 2.5 and 6 kilometers.
3. Clouds found below 2.5 kilometers do not have a specific name.

By its appearance, clouds are classified into three:

1. Strato
   Strato in Latin means layer. Strato clouds are clouds shaped like layers.
2. Cumulo
   Cumulo in Latin means heap. Cumulo clouds are puffy-shaped clouds.
3. Nimbo
   Nimbo in Latin means rain. Nimbo clouds looks a bit similar to Cumulo, but thicker than Cumulo.

Figure 3.1. Source:
http://images.bigcartel.com/bigcartel/product_images/140953480/max_w-2000/CLOUDS.jpg

From those 3 classes, there are 10 common cloud types made:

1. Cirrus
   Cirrus clouds found at high altitude range. The shape of these clouds is threadlike, silky sheen appearance.[7] Cirrus clouds are always composed of ice crystals. The transparency of these clouds depends upon the degree of separation of the ice crystals. These clouds contain no precipitation, meaning they bring no rain.

Figure 3.2. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/ci.jpg

2. Cirrostratus
   Found at high level range, Cirrostratus is translucent cloud veil of ice crystals. This type of clouds sometimes causes halo appearances around moon and sun.[8] Cirrostratus clouds nearly always end by covering the whole sky. Cirrostratus clouds contain no precipitation, meaning it will not make any rainfall.

Figure 3.3. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/cs.jpg

3. Cirrocumulus
   Cirrocumulus composed of ice crystals. The appearance of this cloud type is like snowflakes. These high-level clouds contain no precipitation. In common, Cirrocumulus represents a degraded state of Cirrus and Cirrostratus.

Figure 3.4. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/cc.jpg

4. Altostratus
Altostratus clouds are like dense grey sheet. They only let the sun shines through a little. Found at middle altitude range, Altostratus clouds possibly bring rain or snow.

Figure 3.5. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/as.jpg

5. Altocumulus
Altocumulus is a mid-level cloud. Altocumulus formed a layered plates, compound like rough fleecy cloud. These clouds possibly bring light rain.

Figure 3.6. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/ac.jpg

6. Nimbostratus
Nimbostratus formed from thickening Altostratus clouds. Nimbostratus is a dark grey layer, thick enough to totally block out the sun. This will bring rain or snow, in a heavier intensity than Altostratus.

Figure 3.7. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/ns.jpg

7. Cumulus
Low-level clouds. Cumulus clouds are generally dense with sharp outline whose shape reminds of the form of tower or cotton. Cumulus develops on days of clear skies. It possibly brings rain or snow.

Figure 3.8. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/cu.jpg

8. Stratus
Low level layer if grey cloud. Causes fog and possibly brings drizzle.

Figure 3.9. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/st.jpg

9. Cumulonimbus
Cumulonimbus is a low-level cloud shaped like a real big tower or cauliflower. This is one kind of a thunderstorm cloud, meaning it possibly brings tornadoes and hail. Cumulonimbus clouds also bring rain and snow.

Figure 3.10. Source:
http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/cb.jpg

10. Stratocumulus
Low-level layered cloud which usually colored white or grey. They possibly bring drizzle.

Figure 3.11. Source: http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/images/sc.jpg

From all the descriptions of each cloud types, we can infer some characteristics of clouds: shape, vertical extend, and altitude. The combinations of these characteristics will produce different kind of clouds, the ones who bring rain and the others who do not.

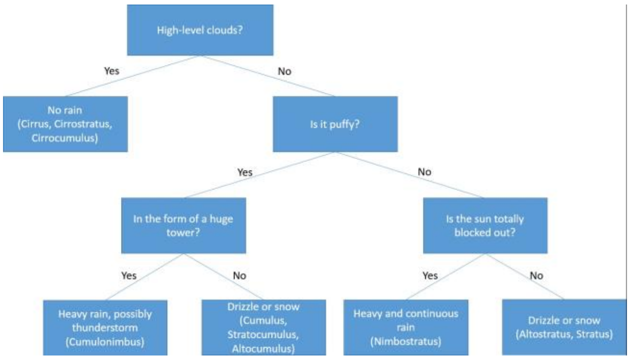


Figure 3.12
Decision Tree of Weather

V. CONCLUSION

There are so many applications of tree, especially decision tree. With decision tree, we can easily solve a problem step by step. I, as writer, suggest the readers to explore more about tree so it can be more applicable in daily life.

VI. ACKNOWLEDGMENT

I would like to thank Allah SWT for His guidance since I started making this paper until it is finally done. I would also express my gratitude to our two lecturers of IF2120 Matematika Diskrit, Dr. Ir. Rinaldi Munir, MT and Dra. Harlili S., M.sc. Thanks to my parents who have been supporting me. Last but not least, thank you to all of the people around me who have accompanied me through my hardest times.

REFERENCES

[1] http://www.ussartf.org/predicting_weather.htm accessed on December 8, 2015 at 7:15 p.m.
[2] Munir, Rinaldi. 2003. *Matematika Diskrit Edisi Kedua*. Bandung: Penerbit Informatika.
[3] Rosen, Kenneth H. 2012. *Discrete Mathematics and Its Applications 7th Edition*. New York: McGraw-Hill.
[4] http://www.atmo.arizona.edu/students/courselinks/fall12/atmo170a1s2/coming_up/week_8/lect22_cloud_identification.html accessed on December 10, 2015 at 4:25 p.m.
[5] http://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/cld/cldtyp/home.rxml accessed on December 10, 2015 at 4:21 p.m.
[6] http://www.bom.gov.au/weather-services/about/cloud/cloud-types.shtml accessed on December 10, 2015 at 4:43 p.m.
[7] http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/types.html accessed on December 10, 2015 at 4.20 p.m.
[8] http://www.clouds-online.com/ accessed on December 10, 2015 at 4:21 p.m.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 10 Desember 2015

Atika Firdaus
13514009