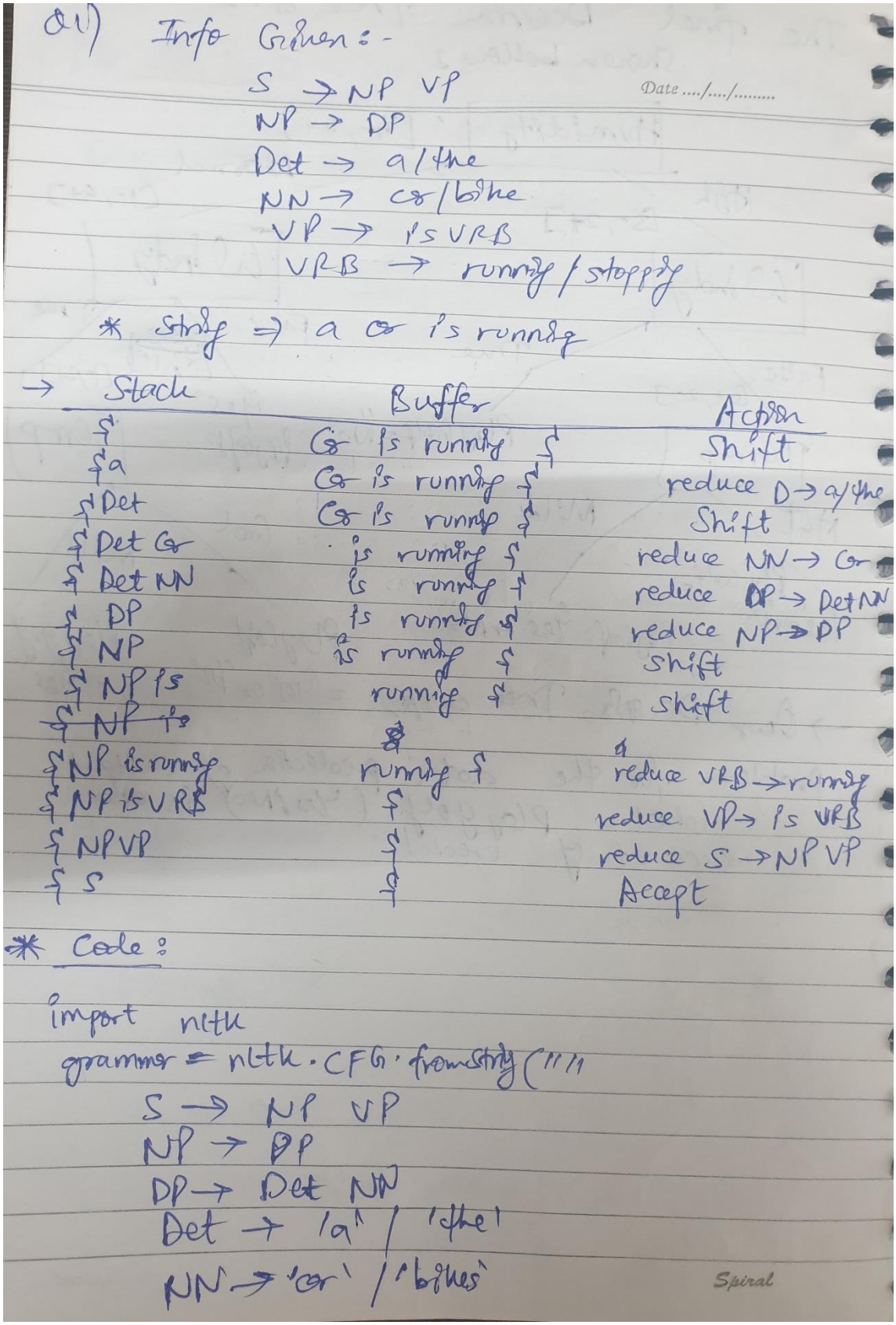
Tanay Mishra 14-04-2023

NLP Dr. Mohamed Iqbal M

Digital Assignment 1



VP → 'is' VRB
VRB → 'running' | 'stopping'

Date ..../..../.....

```
Parser = nltk.ChartParser(grammar)

def generate_strings():
    strings = set()
    for tree in parser.parse(['s']):
        for leaf in tree.leaves():
            strings.add(leaf)
    return strings

acceptable_strings = generate_strings()
print("All possible strings")
for string in acceptable_strings:
    print(string)
```

Output:

```
a car is running
a car is stopping
the car is running
the car is stopping
a bike is running
a bike is stopping
the bike is running
the bike is stopping
```

NLP Digital Assignment 1

Date ...../...../.....

Q2) $P_{yes} = \frac{7}{11}$, $P_{no} = \frac{4}{11}$ Playgolf :- No - 4, Yes - 7

Step 1: For the root node, calculating Entropy of each attribute class :-

Avg Entropy of Dataset (D), Entropy(D) $= \frac{4}{11}\log\frac{4}{11} - \frac{7}{11}\log\frac{7}{11} \Rightarrow 0.946$

Attribute 1 – Temperature

→ It has the following class labels → <Hot, Mild, Cold>

For class label 'Hot', Entropy ↓

Entropy(Temp = "Hot") $\Rightarrow -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$

$\Rightarrow 0.38 + 0.52 \Rightarrow 0.9$

Entropy (Temp = "Mild") $\Rightarrow -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4}$

$\Rightarrow 0.31 + 0.5 \Rightarrow 0.81$

Entropy (Temp = "Cold") $\Rightarrow -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4}$

$\Rightarrow 0.81$

For "Temp" Attribute, we will calculate Infor ↓

$\Rightarrow \frac{3}{14} \times 0.9 + \frac{4}{14} \times 0.81 + \frac{4}{14} \times 0.81$

$\Rightarrow 0.245 + 0.589 \Rightarrow 0.834$

Information gain for "Temp" Attr $\Rightarrow 0.94 - 0.834$ Spiral

$\Rightarrow 0.106$

For Attribute 2 - Humidity

Class Labels → <High, Normal>

Date ..../..../.....

Entropy (Humidity = "High") ↓

$\Rightarrow -\frac{3}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}$

$\Rightarrow 0.52 + 0.44 \Rightarrow \underline{0.96}$

Entropy (Humidity = "Normal") ↓

$\Rightarrow -\frac{5}{6}\log\frac{5}{6} - \frac{1}{6}\log\frac{1}{6}$

$\Rightarrow 0.219 + 0.430 \Rightarrow \underline{0.64}$

For Humidity Attribute, we calc. Information

$\Rightarrow \frac{5}{11} \times 0.96 + \frac{6}{11} \times 0.64$

$\Rightarrow 0.43 + 0.34 \Rightarrow \underline{0.779}$

Information gain for "Humidity" Attr.

$\Rightarrow 0.940 - 0.779$

$\Rightarrow \underline{0.161}$

For Attribute 3 - Windy

Class Labels → <False, True>

Entropy (Windy = "False") ↓

$\Rightarrow -\frac{2}{7}\log\frac{2}{7} - \frac{5}{7}\log\frac{5}{7} \Rightarrow 0.51 + 0.34$

$\Rightarrow \underline{0.85}$

Entropy (Windy = "True") ↓

$\Rightarrow -\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4} \Rightarrow \log 1/2$

$\Rightarrow 1$

Information → $\frac{7}{11} \times 0.85 + \frac{4}{11} \times 1 \Rightarrow 0.54 + 0.36 \Rightarrow \underline{0.903}$

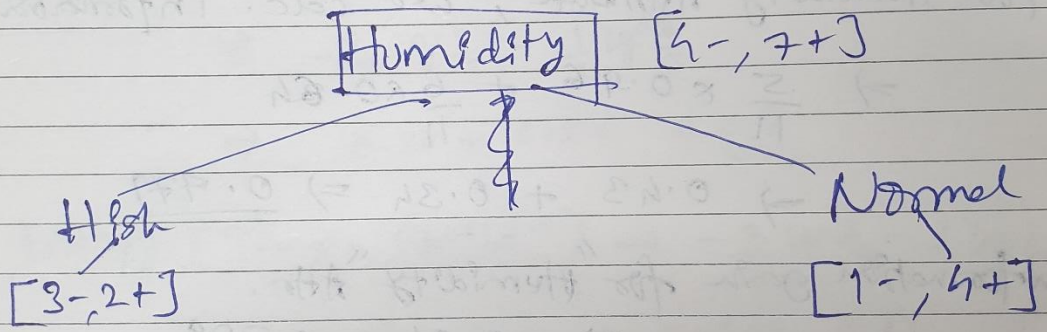
Information gain ⇒ 0.94 - 0.903
→ 0.037

Date ..../..../.....

Step 2 → After getting IG for all Attributes, we compare which one is the maximum and choose it to be the root node.

→ In this case max(IG(Attribute(N))) is "Humidity" with IG = 0.037 being the lowest.

Our Tree so far will look like:



Step 3 → We will do step 1 again but for the root node being "Humidity"

For Attribute = "Temperature" & Humidity = "High":

Entropy < Temp = "Hot", Humidity = "High" >
⇒ $-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$ ⇒ 0.9

Entropy < Temp = "Mild", Humidity = "High" >
⇒ 1

Entropy < Temp = "Cool", Humidity = "High" >
⇒ 0

Spiral

Info → $\frac{3}{5} \times 0.9 + \frac{2}{5} \times 1 + 0$

$\Rightarrow 0.54 + 0.4$

$\Rightarrow 0.94$

Date ..../..../.....

Information Gain $\Rightarrow$ $0.94 - 0.94$

$\Rightarrow 0$

For Attribute = "Windy" & Humidity = "High"

Entropy < Windy = "False" & Humidity = "High" >

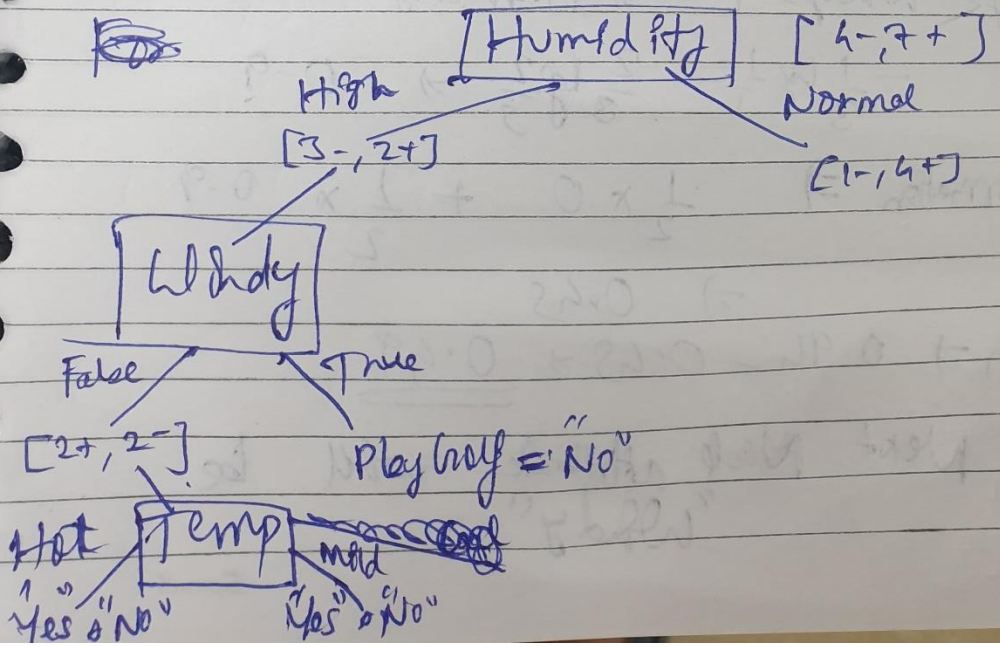
$\Rightarrow -\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4} = 1$

Entropy < Windy = "True" & Humidity = "High" >

$\Rightarrow 0$

Information = $\frac{1}{5} \times 1 + 0 \Rightarrow 0.2$

IG $\Rightarrow$ $0.94 - 0.2 \Rightarrow 0.92$



Spiral

For Attribute = Temp & Humidity = Normal

Date ..../..../.....

Entropy < Temp = Mild, Humidity = Normal >

⇒ $\log 1 \rightarrow 0$

Entropy < Temp = Cool, Humidity = Normal >

⇒ 1

Information ⇒ $0 + \frac{1}{6} \times 1 \Rightarrow 0.66$

IG ⇒ $0.94 - 0.66 \Rightarrow 0.28$

For Attribute = Windy & Humidity = Normal

Entropy < Windy = False, Humidity = Normal >

⇒ 0

Entropy < Windy = True, Humidity = Normal >

⇒ $-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} \Rightarrow 0.9$

Information ⇒ $\frac{1}{2} \times 0 + \frac{1}{2} \times 0.9$

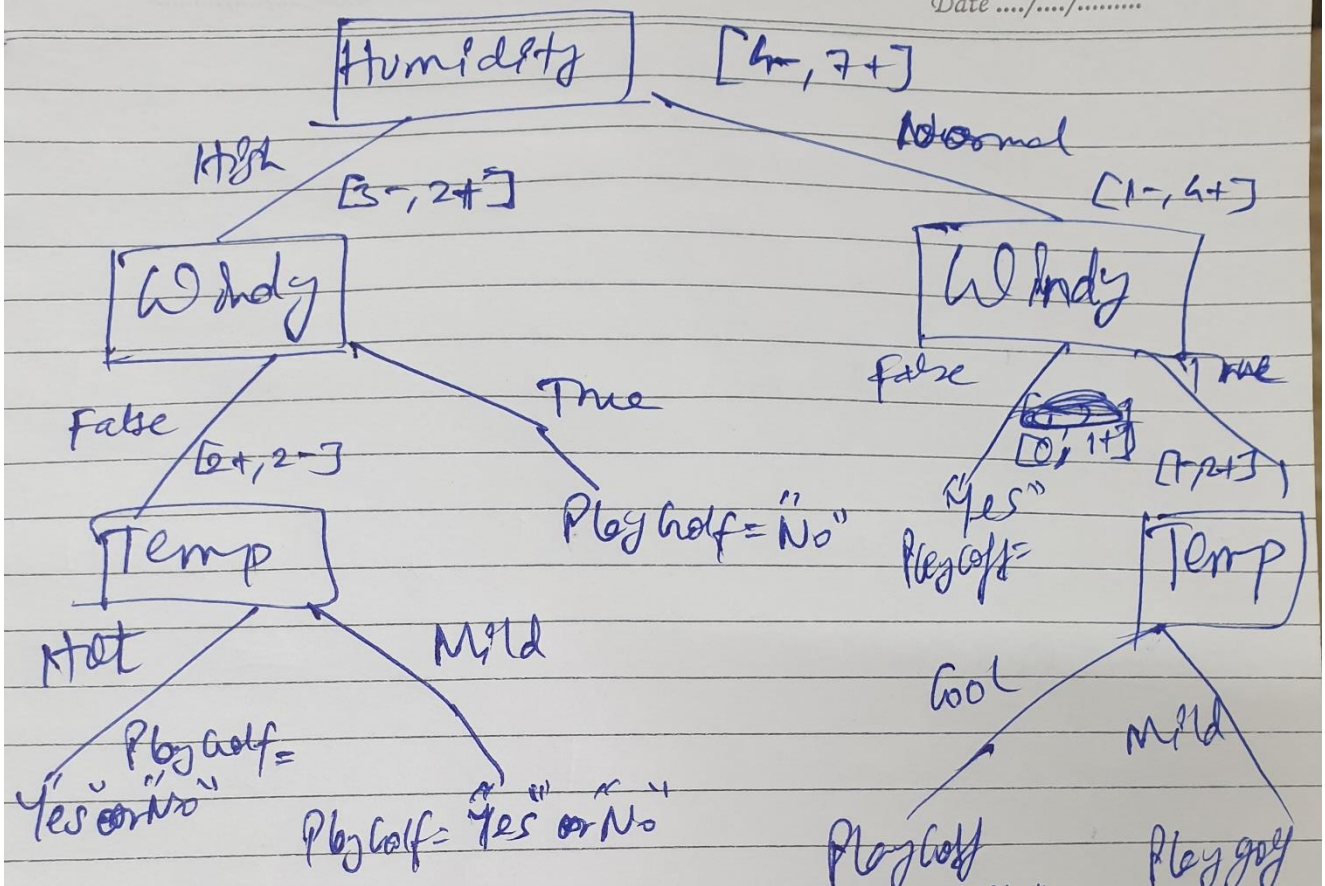
⇒ 0.45

IG ⇒ $0.94 - 0.45 \Rightarrow 0.49$

∴ Our Next Node attribute will be "Windy"

The final Decision Tree is as shown bellow :

Date ..../..../.....



→ Our Decision Tree classifier problem for the data prediction of output label = Play golf ("Yes/No") is successfully created.

Q3) Given Information:-

S → NP VP
NP → Det NN
Det → a/an/the
NN → Child/Adult
VP → AUX . VRB
AUX → is/was
VRB → cryingf/sleepingf

Date ..../..../.....

* Given sentence a child is cryingf ↓

Stack	Input Buffer	Action
$	A child is cryingf $	Shift
$ a	child is cryingf $	reduce Det → a
$ Det	child is cryingf $	Shift
$ Det child	is cryingf $	reduce NN → child
$ Det NN	is cryingf $	reduce NP → Det NN
$ NP	is cryingf $	Shift
$ NP is	cryingf $	reduce AUX → is
$ NP AUX	cryingf $	shift
$ NP AUX cryingf	$	reduce VRB → cryingf
$ NP AUX VRB	$	reduce VP → AUX VRB
$ NP VP	$	reduce S → NP VP
$ S	$	accept

* Code

```
import nltk
grammar = nltk.CFG.fromstring("""
    S → NP VP
    NP → Det NN
    Det → 'a' | 'an' | 'the'
    NN → 'child' | 'adult'
    VP → AUX VRB
    AUX → 'is' | 'was'
```

Spiral

```
VRB -> 'cryng' / 'sleeping' | ....)
Parser = nltk.ChartParser(grammar)/.....
def generate():
    strng = set()
    for tree in parser.parse(''.join(list)):
        for leaf in tree.leaves():
            strng.add(leaf)
    return strng

as = generate_strng():
for s in as:
    print(s)
```

* Output

A child is crying
An child is crying
the child is crying
~~A child is crying~~
A adult is crying
An adult is crying
the adult is crying
A child was crying
An child was crying
the child was crying
A adult was crying
An adult was crying
the adult was crying.

A child is sleeping
An child is sleeping
the child is sleeping
A adult is sleeping
An adult is sleeping
the adult is sleeping
A child was sleeping
An child was sleeping
the child was sleeping
A adult was sleeping
an adult was sleeping
the adult was sleeping

Q4)

$P(\text{Buy Car} = \text{Yes}) = 6/10 = 0.6$ Date ..../..../.....

$P(\text{Buy Car} = \text{No}) = 4/10 = 0.4$

Family type	Y	N
Nuclear	3/6	0
Extended	1/6	1/4
Childless	1/6	2/4
Single Parent	1/6	1/4

Age Group	Y	N
~~Young~~	~~2/6~~	—
Young	2/6	1/4
Middle Age	3/6	1/4
Old	1/6	2/4

Income Status	Y	N
Low	1/6	4/4
Medium	3/6	0
High	2/6	0

To find for :-

Single parent, young, high

$V_{NB}(\text{yes}) = (0.6)(1/6)(2/6)(2/6) = 0.011$

$V_{NB}(\text{No}) = (0.4)(1/4)(1/4)(0) = 0$

∴ Yes, they will purchase the car

Q5) Given Info :-

S → NP VP

NP → DP

DP → Det NN

Det → a / the

NN → car / bike

VP → is VRB

VRB → running / stopping

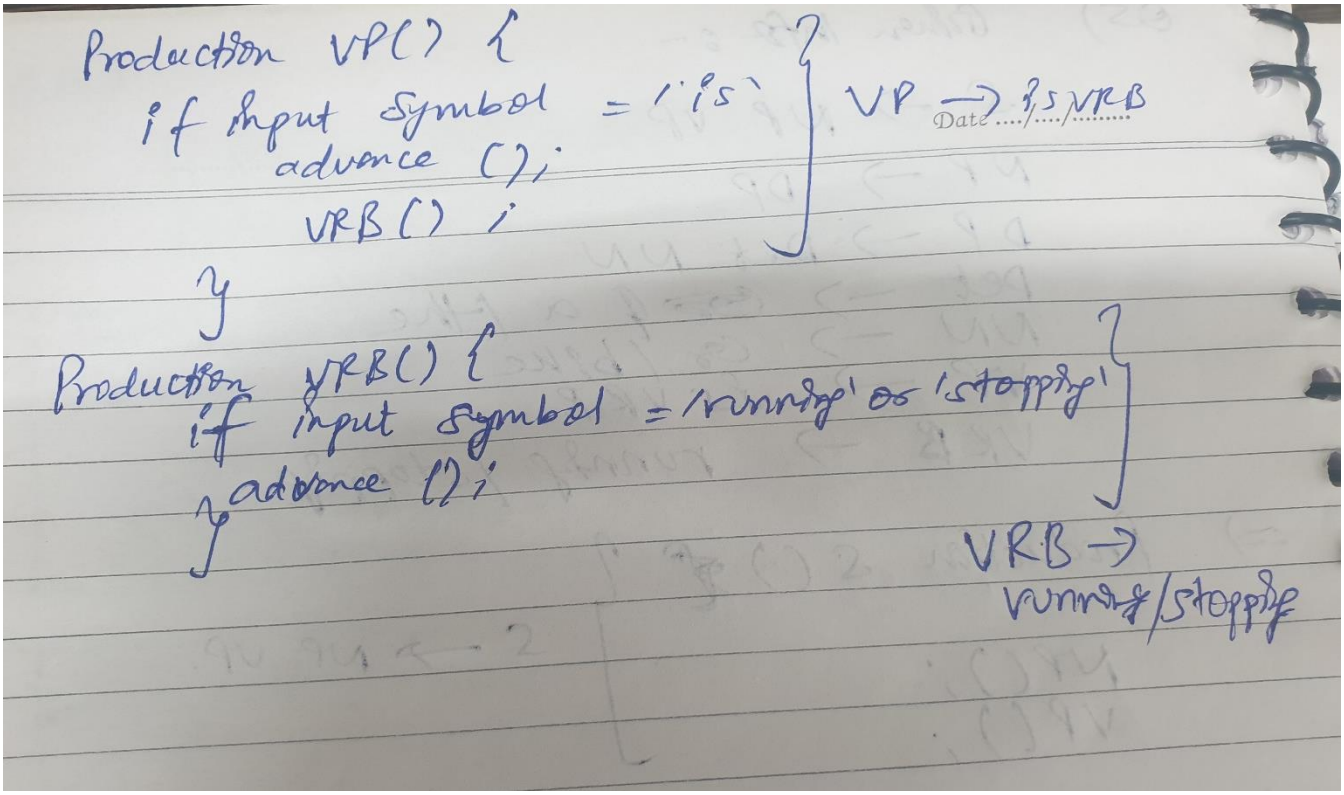
⇒ Production S() {
    NP();
    VP();
} S → NP VP.

Production NP() {
    DP();
} NP → PP

Production DP() {
    Det();
    NN();
} DP → Det NN

Production Det() {
    If Input symbols = 'a' or 'the'
        advance();
} Det → a/the

Production NN() {
    If Input symbol = 'car'
        advance();
    else If Input symbol = 'bike'
        advance();
} NN → car/bike.


Production VP() {
if input symbol = 'is'
advance();
VRB();
}
VP → is VRB
Production VRB() {
if input symbol = 'running' or 'stopping'
advance();
}
VRB →
running/stopping