Using C4.5, find the first braching factor for dataset given below.

X_1	X_2	X_3	X_4	D
K	P	G	5	H
K	N	G	1	H
B	E	G	3	H
B	P	G	5	S
B	N	T	-1	S
K	E	T	2	S

1. Compute independent class entropy

   $H(D) = -(0.5*\log_2 0.5+0.5*\log_2 0.5)=1$

2. Make unique and sort X_4 values at first.

X_4	-1	1	2	3	5

3. Determine all thresholds for X_4

$t_1=(-1+1)/2=0$ $t_2=(1+2)/2=1.5$ $t_3=(2+3)/2=2.5$ $t_4=(3+5)/2=4$

4. By using each threshold, prepare the dataset again. Then compute entropy for each new table. According to information gain, choose the best dataset (so threshold).

for t_1=0	
X_4	D
A	H
A	H
A	H
A	S
B	S
A	S

for t_2=1.5	
X_4	D
A	H
B	H
A	H
A	S
B	S
A	S

for t_3=2.5	
X_4	D
A	H
B	H
A	H
A	S
B	S
B	S

for t_4=4	
X_4	D
A	H
B	H
B	H
A	S
B	S
B	S

$H(X_4t_1|D) = 5/6*H(A)+1/6*H(B) = 5/6(-(0.6*\log_2 0.6+0.4*\log_2 0.4))+1/6*0 \approx$ **0.81**
$H(X_4t_2|D) = 4/6*H(A)+2/6*H(B) = 4/6*1+2/6*1 =$ **1**
$H(X_4t_3|D) = 3/6*H(A)+3/6*H(B) = 3/6*(-(2/3*\log_2 2/3+1/3*\log_2 1/3))+ 3/6*(-(1/3*\log_2 1/3+2/3*\log_2 2/3)) \approx$ **0.92**
$H(X_4t_4|D) = 2/6*H(A)+4/6*H(B) =$ **1**

Because the maximum information gain is obtained from the first threshold (t_1), choose it as the threshold.

5. Compute the entropy for all other categorical features in the table. According to information gain, choose the best feature.

$H(X_1|D) = 1/2*H(K)+1/2H(B) \approx$ **0.92**

$H(X_2|D) = 1/3*H(P)+1/3*H(N)+1/3*H(E) =$ **1**

$H(X_3|D) = 4/6*H(G)+2/6*H(T) = 4/6(-(0.25*\log_2 0.25+0.75*\log_2 0.75))+2/6*0 \approx$ **0.54**

6. Calculate information gain for each feature

$IG(X_1) = H(D) – H(X_1|D) = 1 – 0.92 = 0.08$
$IG(X_2) = H(D) – H(X_2|D) = 1 – 1 = 0$
$IG(X_3) = H(D) – H(X_3|D) = 1 – 0.54 = 0.46$
$IG(X_4) = H(D) – H(X_4|D) = 1 – 0.81 = 0.19$

According to the maximum information gain, we choose the third feature (**X_3**) for the first branching.