|  |  |  |  |
| --- | --- | --- | --- |
| **Past Trend** | **Open Interest** | **Trading Volume** | **Return** |
| Positive | Low | High | Up |
| Negative | High | Low | Down |
| Positive | Low | High | Up |
| Positive | High | High | Up |
| Negative | Low | High | Down |
| Positive | Low | Low | Down |
| Negative | High | High | Down |
| Negative | Low | High | Down |
| Positive | Low | Low | Down |
| Positive | High | High | Up |



Gini for Past Trend = Weighted sum [ Gini (Past trend = Positive) and Gini(Past Trend = Negative)]

Gini (Past trend = Positive) = 1 – sum(Pi)^2 where i is classes

= 1 – (4/6)^2 – (2/6)^2 = 0.444

Gini (Past trend = Negative) = 1 – sum(Pi)^2 where i is classes

= 1 – 0 – (4/4)^2 = 0

Gini for Past Trend = (6/10)\*0.444 + (4/10)\*0 = 0.2666

Gini for Open Interest = Weighted sum of Gini (Open Interest = High) and Gini(Open Interest = Low)

Gini (Open Interest = High) = 1 – sum(Pi)^2 where i is classes

= 1 – (2/4)^2 – (2/4)^2 = 0.5

Gini (Open Interest = Low) = 1 – sum(Pi)^2 where i is classes

= 1 – (2/6)^2 – (4/6)^2 = 0.444

Gini for Open Interest = (4/10)\*0.5 + (6/10)\*0.444 = 0.466

|  |  |  |  |
| --- | --- | --- | --- |
| **Past Trend** | **Open Interest** | **Trading Volume** | **Return** |
| Positive | Low | High | Up |
| Negative | High | Low | Down |
| Positive | Low | High | Up |
| Positive | High | High | Up |
| Negative | Low | High | Down |
| Positive | Low | Low | Down |
| Negative | High | High | Down |
| Negative | Low | High | Down |
| Positive | Low | Low | Down |
| Positive | High | High | Up |



Gini for Trading Volume = Weighted sum of Gini (Trading volume = High) and Gini(Trading volume= Low)

Gini (Trading Volume = High) = 1 – sum(Pi)^2 where i is classes

= 1 – (4/7)^2 – (3/7)^2 = 0.489

Gini (Trading Volume = Low) = 1 – sum(Pi)^2 where i is classes

= 1 – (0) – (3/3)^2 = 0

Gini for Trading Volume = (7/10)\*0.489 + (3/10)\*0 = 0.3423

Gini : Past Trend = 0.26, Opn Intrest=0.466 and Trading volume = 0.3423

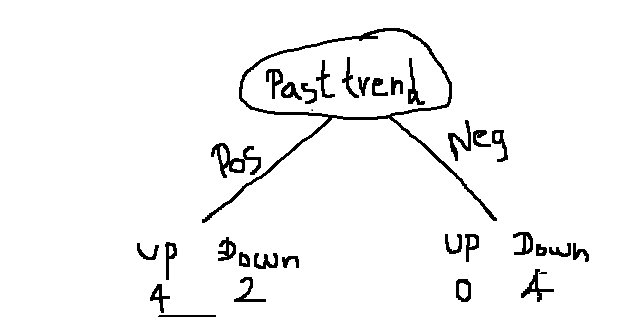
Gini Target = 1- (4/10)^2+(6/10)^2 = 0.48

Gini Gain : G(Target) – G(X)

Gini gain : Past Trend = 0.48-0.26 = 0.22, Opn Intrs = 0.14 Trading vol = 0.14

Features Selected is based on feature with low gini or high gini gain

**Keep Past Trend as first root node as its gini index is less or Gini gain is high and split the data**

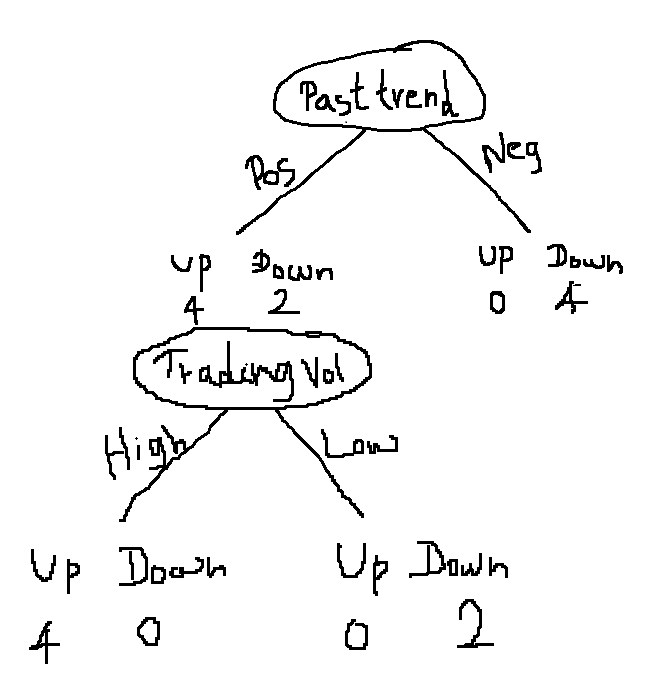


|  |  |  |  |
| --- | --- | --- | --- |
| **Past Trend** | **Open Interest** | **Trading Volume** | **Return** |
| Positive | Low | High | Up |
| Positive | Low | High | Up |
| Positive | High | High | Up |
| Positive | Low | Low | Down |
| Positive | Low | Low | Down |
| Positive | High | High | Up |

Gini Index for Open Interest = 0.33

Gini for Trading volume = 0

**Keep Trading volume as split node as its gini index is less and split the data**

****

|  |  |  |  |
| --- | --- | --- | --- |
| **Past Trend** | **Open Interest** | **Trading Volume** | **Return** |
| Positive | Low | High | Up |
| Negative | High | Low | Down |
| Positive | Low | High | Up |
| Positive | High | High | Up |
| Negative | Low | High | Down |
| Positive | Low | Low | Down |
| Negative | High | High | Down |
| Negative | Low | High | Down |
| Positive | Low | Low | Down |
| Positive | High | High | Up |

**Information Gain(Past Trend) = E(Return) – E(Past trend | Return)**

E(return) = sum (-pi\*logpi)

E(return) = -(6/10)\*log(6/10) – (4/10)\*log(4/10) = 0.44 + 0.528 = 0.97

E(past trend = Positive | Return) = -(4/6)\*log(4/6) – (2/6)\*log(2/6) = 0.3899 + 0.528 = 0.918

E(past trend = Negative | Return) = -(0/4)\*log(0/4) – (4/4)\*log(4/4) = 0

E(past trend | Return) = 6/10 \* 0.918 + 4/10\*0 = 0.5508

IG = 0.97 – 0.5508 = 0.4192

Conclusion:

1. Column with low gini or high gini gain is selected as the node.
2. Column with low entropy or high information gain is selected as the node.

**GINI for Continuous feature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Past Trend** | **Open Interest** | **Trading Volume** | **Return** |  |
| Positive | Low | 280 | Down |  |
| Negative | High | 300 | Down | 290 |
| Positive | Low | 320 | Down | 310 |
| Positive | Low | 340 | Up | 330 |
| Positive | Low | 360 | Up | **350** |
| Positive | High | 565 | Up | 462 |
| Negative | Low | 652 | Down | 608 |
| Negative | Low | 695 | Down | 673 |
| Negative | High | 715 | Down | 705 |
| Positive | High | 760 | Up | 737 |

**Gini for Trading volume value 350 = weighted sum ( gini for trading volume above 350 + gini for trading volume below 350)**

Gini (trade vol > 350) = 1 – (3/6)^2 –(3/6)^2 = 0.5

Gini (trade vol < 350) = 1 – (1/4)^2 –(3/4)^2 = 0.375

Gini for Trading volume value 350 = (6/10)\*0.5 + (4/10)\*0.375 = 0.45

Similarly compute for all the moving average values of Trading volume and see for which value gini impurity is less. Use that value for splitting the trading volume.