1. (a)

$E(Z) = E(X^2 - 2XY + Y^2) = E(X^2) - 2E(XY) + E(Y^2) = E(X^2) - 2E(X)E(Y) + E(Y^2)$

$= \frac{1}{6}$

for uniform distribution

$E(X) = \frac{1}{2}(a+b) = \frac{1}{2}$

$V(X) = \frac{1}{12}(b-a)^2 = \frac{1}{12}$

$VAR(X) = E(X^2) - E(X)^2$

$E(X^2) = \frac{1}{3}$

$VAR(Z) = E(Z^2) - E(Z)^2$

$= E((X-Y)^4) - E(Z)^2$

$E(X^4) = \int_0^1 x^4 \cdot 1\,dx = \frac{1}{5} = E(Y^4)$

$E(X^3) = \int_0^1 x^3 \cdot 1\,dx = \frac{1}{4} = E(Y^3)$

$VAR(Z) = E(X^4) - 4E(X^3)E(Y) - 4E(X)\cdot E(Y^3) + 6E(X^2)E(Y^2) + E(Y^4) - E(Z)^2$

$= \frac{7}{180}$

1. (b) $E(R) = E(Z_1 + Z_2 \cdots Z_d) = E(Z_1) + E(Z_2) + \cdots E(Z_d) = \frac{d}{6}$

$V(R) = V(Z_1 + Z_2 \cdots Z_d) = V(Z_1) + V(Z_2) + \cdots V(Z_d) + 2cov(Z_1, Z_2) + 2cov(Z_{n-1}, Z_n)$

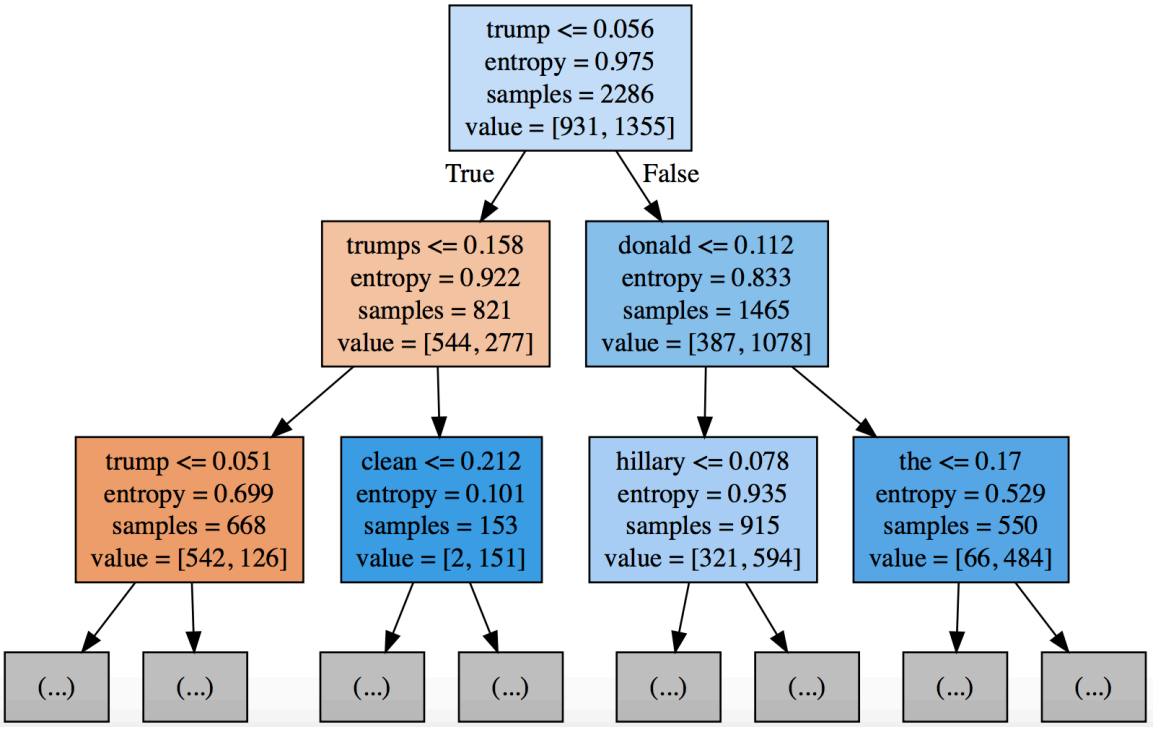
since Z_i are independent, there is no covariance

so $V(R) = \frac{7}{180} d$

2(b)

```
mode is gini
5
Validation Accuracy: 0.8
mode is gini
50
Validation Accuracy: 0.7918367346938775
mode is gini
150
Validation Accuracy: 0.8061224489795918
mode is gini
250
Validation Accuracy: 0.789795918367347
mode is gini
400
Validation Accuracy: 0.7938775510204081
mode is entropy
5
Validation Accuracy: 0.810204081632653
mode is entropy
50
Validation Accuracy: 0.8040816326530612
mode is entropy
150
Validation Accuracy: 0.8081632653061225
mode is entropy
250
Validation Accuracy: 0.8163265306122449
mode is entropy
400
Validation Accuracy: 0.8081632653061225
```

2(c)



2(d)

```
118         Info_Gain = HY - HY_X
119         print(Info_Gain)
120         return Info_Gain
121     # select_model()
122     compute_information_gain("trumps")
123
124
```

```
/Users/sheepinwolfskin/anaconda3/env
0.044381813216067045

Process finished with exit code 0
```

```
    # select_model()
    compute_information_gain("hillary")
```

```
/Users/sheepinwolfskin/anaconda3/env
0.03634944051710376

Process finished with exit code 0
```