1. Section A

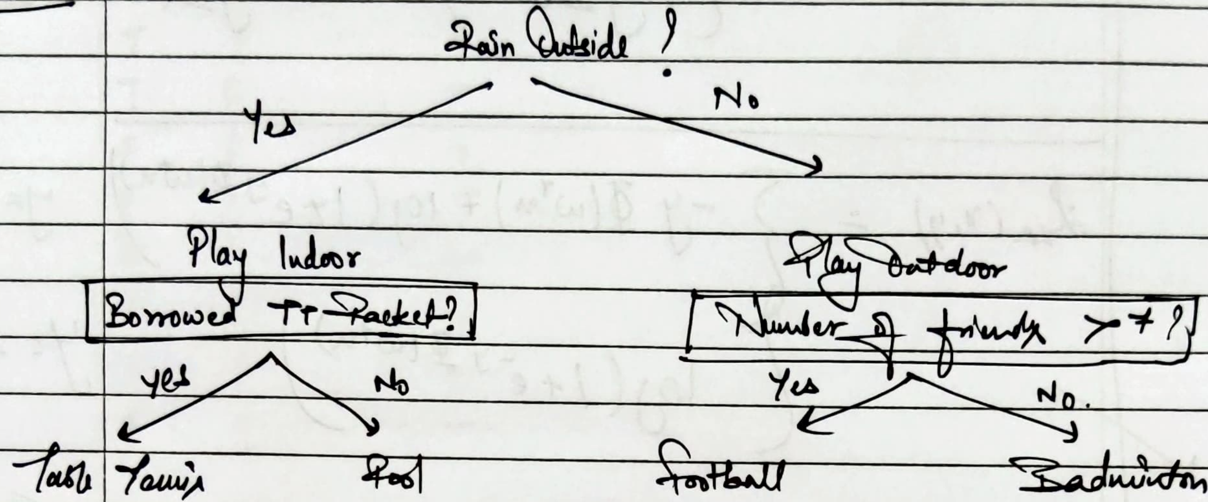
a. While a certain level of correlation among trees is desirable to maintain ensemble stability and predictive power, excessive correlation can lead to overfitting and reduced diversity, limiting the model's generalization capabilities. Striking the right balance is essential for a robust Random Forest model that leverages both individual tree strengths and the power of ensemble diversity.

b. The "curse of dimensionality" can become an issue in Naive Bayes when the number of features (dimensions) in the dataset is large, leading to increased computational complexity. Dimensionality reduction techniques such as feature selection or extraction can help mitigate this problem in practice.

c. If the Naive Bayes classifier encounters a value of attributes not present in the training dataset, it may face issues because it assumes conditional independence between features given the class label. This can lead to biased or unreliable predictions. To mitigate this problem, approaches like Laplace smoothing (add-one smoothing) can be used to assign a small, non-zero probability to unseen values, making the classifier more robust. Alternatively, encoding unseen values as a special category or using more advanced models like tree-based or neural network classifiers can also help address this issue. For example, in a spam email classification task, encountering new and previously unseen words in emails can lead to misclassifications, so using Laplace smoothing can help account for these situations.

d. Yes, Information Gain can be biased when attributes have different cardinalities because it tends to favor attributes with more distinct values. To address this bias, Gini Impurity is another criterion for attribute selection that is less sensitive to cardinality. Gini Impurity measures the probability of misclassifying a randomly chosen element if it were classified according to the class distribution in a given set. It doesn't explicitly favor attributes with high cardinality and can be a more balanced criterion. For example, in a decision tree for customer churn prediction, if you have a binary attribute like "HasCreditCard" with low cardinality and another like "TransactionHistory" with high cardinality, Gini Impurity can provide a more fair evaluation of these attributes for splitting decisions.

Ques - 2) a.)



→ P(TT) = P(TT racket / raining) * P(raining)

→ P(Pool) = P(no TT racket / raining) * P(raining)

→ P(football) = P(> 7 friends / not raining) * P(not raining)

→ P(Badminton) = P(< 7 friends / not raining) * P(not raining).

b.) Using Bayes theorem

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

A → going to rain
B → App predicts rainy.

P(rain) = 0.3 P(Predicted rainy / rain) = 0.8
P(clear) = 0.7 P(Predicted clear / clear) = 0.9

using law of total prob.

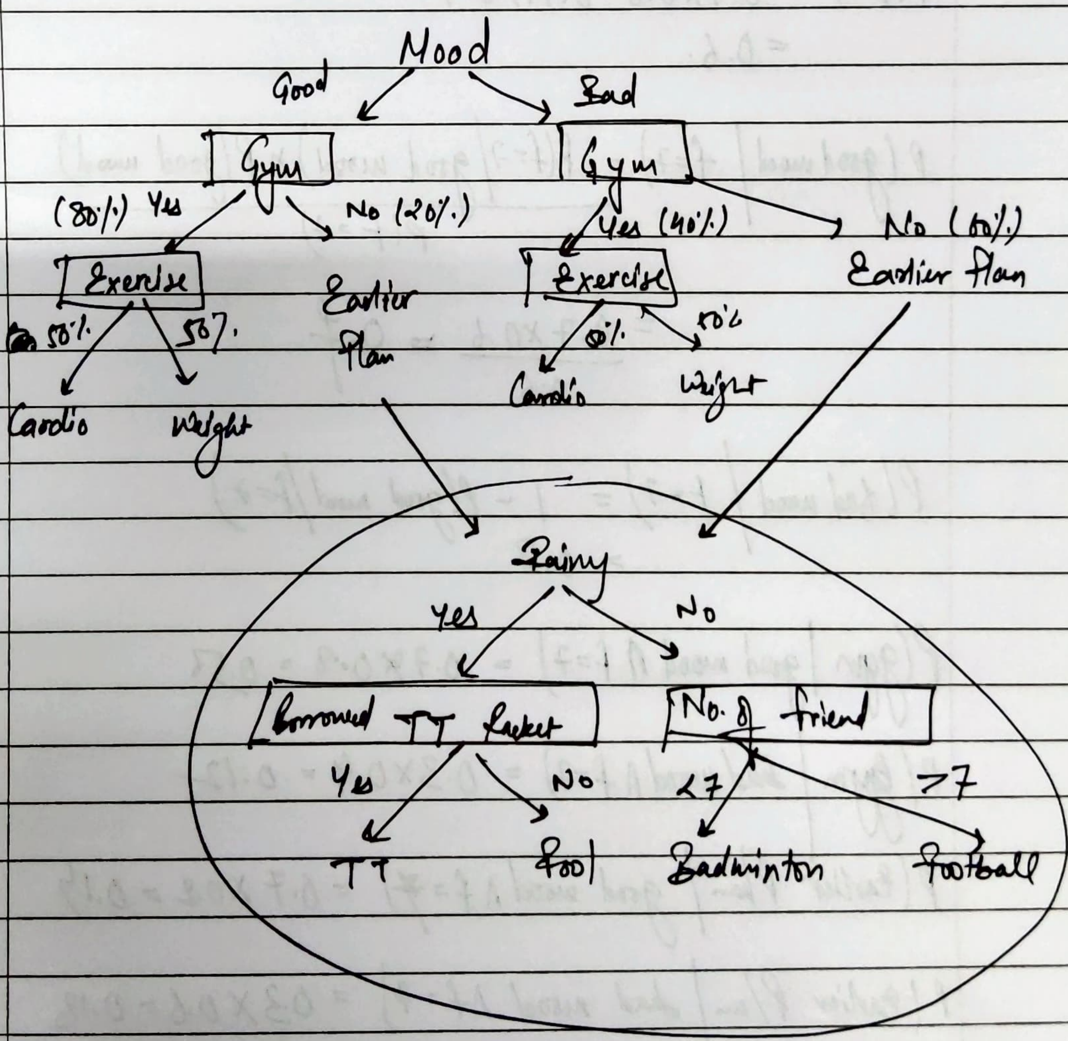
P(Predicted rainy) = 0.7 × 0.1 + 0.8 × 0.3
= 0.31

Now, P(Rain / Predicted Rainy) = $\frac{P(\text{Predicted Rainy} / \text{Rain}) \times P(\text{rain})}{P(\text{Predicted Rainy})}$

$$= \frac{0.8 \times 0.3}{0.31}$$

$$= \underline{0.77419.}$$

c) P(Gym | Good Mood) = 0.8
P(earlier plan | Good Mood) = 0.2
P(gym | Bad mood) = 0.4
P(earlier plan | Bad mood) = 0.6



d) Given, $P(\text{good mood}) = 0.6$.
$P(\text{bad mood}) = 0.4$

$P(f=7 \mid \text{good mood}) = 0.7$.

$P(f=7 \mid \text{Bad mood}) = 0.45$

$P(f=7) = 0.7 \times 0.6 + 0.45 \times 0.4$
$= 0.6$.

$$P(\text{good mood} \mid f=7) = \frac{P(f=7 \mid \text{good mood}) \times P(\text{good mood})}{P(f=7)}$$

$$= \frac{0.7 \times 0.6}{0.6} = 0.7$$

$P(\text{bad mood} \mid f=7) = 1 - P(\text{good mood} \mid f=7)$
$= 0.3$

$P(\text{gym} \mid \text{good mood} \wedge f=7) = 0.7 \times 0.8 = 0.56$

$P(\text{gym} \mid \text{bad mood} \wedge f=7) = 0.3 \times 0.4 = 0.12$

$P(\text{Earlier Plan} \mid \text{good mood} \wedge f=7) = 0.7 \times 0.2 = 0.14$

$P(\text{Earlier Plan} \mid \text{bad mood} \wedge f=7) = 0.3 \times 0.6 = 0.18$

$P(\text{gym} \mid f=7) = 0.56 + 0.12$
$= 0.68$

$P(\text{go Earlier plan} \mid f=7) = 0.14 + 0.18$
$= 0.32$

$\therefore$ He will most likely go gym