**Question 1:**

You have been presented with a dataset from a major bank containing information about 15 customers. The dataset consists of three independent variables describing the financial situation of the customers (Credit Score (numeric); Debt-to-Income Ratio (numeric); Number of Late Payments (numeric)) and one dependent variable indicating whether a customer failed to pay their loan (Loan Default (binary: 1 for default, 0 for no default)).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Customer** | **Credit**  **Score** | **Debt-to-**  **Income Ratio** | **Number of Late**  **Payments** | **Loan Default** |
| 1 | 700 | 0.3 | 1 | 0 |
| 2 | 650 | 0.4 | 2 | 1 |
| 3 | 600 | 0.5 | 3 | 1 |
| 4 | 720 | 0.2 | 0 | 0 |
| 5 | 680 | 0.3 | 2 | 0 |
| 6 | 630 | 0.6 | 3 | 1 |
| 7 | 690 | 0.4 | 1 | 0 |
| 8 | 710 | 0.3 | 0 | 0 |
| 9 | 660 | 0.5 | 2 | 1 |
| 10 | 640 | 0.4 | 1 | 1 |
| 11 | 680 | 0.3 | 0 | 0 |
| 12 | 730 | 0.2 | 1 | 0 |
| 13 | 620 | 0.6 | 2 | 1 |
| 14 | 670 | 0.4 | 3 | 1 |
| 15 | 700 | 0.3 | 1 | 0 |

1. Develop a decision tree model to predict credit default. Interpret your result.
2. Suggest policies that can be made by the bank to reduce the risk of credit default.

**Question 2:**

A transportation research institute is studying traffic accidents to identify factors that contribute to the severity of accidents. They have collected data on 15 accidents and want to use decision tree modeling to predict the severity of accidents based on various factors.

# Data:

The dataset, accident\_data.csv, includes the following variables:

* accident\_id: Unique identifier for each accident
* speed\_limit: Speed limit at the accident location (km/h)
* weather\_conditions: Weather conditions during the accident (e.g., sunny, rainy, snowy)
* road\_type: Type of road (e.g., highway, urban street, rural road)
* time\_of\_day: Time of day the accident occurred
* severity: Severity of the accident (e.g., minor, moderate, severe)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| accident\_id | speed\_limit | weather\_conditions | road\_type | time\_of\_day | severity |
| 1 | 50 | Sunny | Highway | Day | Minor |
| 2 | 60 | Rainy | Urban Street | Night | Moderate |
| 3 | 40 | Snowy | Rural Road | Day | Severe |
| 4 | 70 | Sunny | Highway | Night | Severe |
| 5 | 50 | Rainy | Urban Street | Day | Minor |
| 6 | 60 | Snowy | Rural Road | Night | Moderate |
| 7 | 40 | Sunny | Highway | Day | Minor |
| 8 | 70 | Rainy | Urban Street | Night | Severe |
| 9 | 50 | Snowy | Rural Road | Day | Moderate |
| 10 | 60 | Sunny | Highway | Night | Minor |
| 11 | 40 | Rainy | Urban Street | Day | Severe |
| 12 | 70 | Snowy | Rural Road | Night | Moderate |
| 13 | 50 | Sunny | Highway | Day | Minor |
| 14 | 60 | Rainy | Urban Street | Night | Severe |
| 15 | 40 | Snowy | Rural Road | Day | Moderate |

Please generate a decision tree model using this data, show and explain the R or Python code you are using, and interpret the result. Based on the decision tree and model evaluation results, propose two potential policy recommendations to improve road safety.

Question 3:

You have been provided with a dataset from a manufacturing company containing information about 15 products. The dataset consists of three independent variables describing the production process (Machine Temperature, Production Speed, Raw Material Quality) and one dependent variable indicating whether a product is defective (Defective Product = 0; non defective = 1).

*Table 4*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product** | **Machine Temperature** | **Production Speed** | **Raw Material Quality** | **Defective Product (dummy variable)** |
| 1 | 180 | 100 | 0.7 | 0 |
| 2 | 160 | 90 | 0.6 | 1 |
| 3 | 170 | 95 | 0.8 | 1 |
| 4 | 185 | 105 | 0.5 | 0 |
| 5 | 175 | 100 | 0.6 | 0 |
| 6 | 165 | 85 | 0.8 | 1 |
| 7 | 190 | 110 | 0.7 | 0 |
| 8 | 180 | 100 | 0.6 | 0 |
| 9 | 170 | 95 | 0.5 | 1 |
| 10 | 165 | 90 | 0.7 | 1 |
| 11 | 175 | 100 | 0.6 | 0 |
| 12 | 185 | 105 | 0.8 | 0 |
| 13 | 168 | 93 | 0.7 | 1 |
| 14 | 178 | 97 | 0.6 | 1 |
| 15 | 182 | 103 | 0.5 | 0 |

1. Develop a decision tree model to predict defective products using the given dataset. Interpret the result and discuss the key factors influencing product defects. Show your work using screenshot.
2. Propose policies that the manufacturing company can implement to reduce the occurrence of defective products based on the insights from the decision tree analysis.