**Problem 1: Linear Regression:**

Problem: Predict the house prices based on various features.

Dataset:

| **Square Footage** | **Bedrooms** | **Bathrooms** | **Garage Size** | **Price ($)** |
| --- | --- | --- | --- | --- |
| 2000 | 3 | 2 | 2 | 300,000 |
| 1600 | 2 | 1 | 1 | 240,000 |
| 3000 | 4 | 3 | 2 | 450,000 |
| 2200 | 3 | 2 | 2 | 320,000 |
| 1800 | 2 | 1 | 1 | 260,000 |
| 2500 | 4 | 2 | 3 | 380,000 |
| 2800 | 3 | 3 | 2 | 390,000 |
| 1900 | 2 | 2 | 1 | 270,000 |
| 2100 | 3 | 1 | 2 | 310,000 |
| 1700 | 2 | 2 | 1 | 250,000 |
| 2600 | 4 | 2 | 3 | 360,000 |
| 2300 | 3 | 2 | 2 | 330,000 |
| 2700 | 4 | 3 | 3 | 400,000 |
| 2400 | 3 | 2 | 2 | 340,000 |
| 2900 | 4 | 2 | 3 | 410,000 |
| 3000 | 3 | 3 | 2 | 420,000 |
| 1850 | 2 | 1 | 1 | 255,000 |
| 3200 | 4 | 3 | 3 | 430,000 |
| 2800 | 3 | 2 | 2 | 385,000 |
| 3100 | 4 | 2 | 3 | 440,000 |
| 1800 | 3 | 1 | 2 | 275,000 |
| 2700 | 2 | 2 | 1 | 350,000 |
| 3300 | 4 | 3 | 3 | 470,000 |
| 2600 | 3 | 2 | 2 | 365,000 |
| 3000 | 4 | 2 | 3 | 430,000 |
| 2500 | 3 | 2 | 2 | 330,000 |
| 3500 | 4 | 3 | 3 | 490,000 |
| 2900 | 3 | 2 | 2 | 395,000 |
| 3200 | 4 | 3 | 3 | 480,000 |
| 3000 | 3 | 2 | 2 | 400,000 |

**Problem 2: Logistic Regression:**

Problem: Predict whether a transaction is fraudulent or not based on transaction details.

Dataset:

| **Transaction ID** | **Amount ($)** | **Merchant** | **Time** | **Fraudulent** |
| --- | --- | --- | --- | --- |
| 1 | 1000 | Online Shop | 10:15 AM | No |
| 2 | 150 | Gas Station | 3:30 PM | Yes |
| 3 | 3000 | Retail Store | 8:45 AM | No |
| 4 | 200 | Online Shop | 12:00 PM | No |
| 5 | 500 | Gas Station | 6:20 PM | Yes |
| 6 | 1200 | Online Shop | 9:30 AM | No |
| 7 | 800 | Retail Store | 2:00 PM | No |
| 8 | 300 | Gas Station | 4:45 PM | Yes |
| 9 | 600 | Online Shop | 11:30 AM | No |
| 10 | 2000 | Gas Station | 5:10 PM | Yes |
| 11 | 400 | Retail Store | 7:00 AM | No |
| 12 | 900 | Online Shop | 1:20 PM | No |
| 13 | 700 | Gas Station | 3:50 PM | Yes |
| 14 | 1100 | Retail Store | 10:45 AM | No |
| 15 | 350 | Online Shop | 12:30 PM | No |
| 16 | 250 | Gas Station | 5:40 PM | Yes |
| 17 | 1800 | Online Shop | 8:00 AM | No |
| 18 | 950 | Retail Store | 2:50 PM | No |
| 19 | 500 | Gas Station | 4:15 PM | Yes |
| 20 | 1300 | Online Shop | 11:00 AM | No |
| 21 | 300 | Retail Store | 6:30 AM | No |
| 22 | 600 | Gas Station | 1:50 PM | Yes |
| 23 | 400 | Online Shop | 3:20 PM | No |
| 24 | 700 | Retail Store | 10:00 AM | No |
| 25 | 550 | Gas Station | 12:40 PM | Yes |
| 26 | 1500 | Online Shop | 4:00 PM | No |
| 27 | 900 | Retail Store | 9:15 AM | No |
| 28 | 800 | Gas Station | 2:30 PM | Yes |
| 29 | 200 | Online Shop | 7:50 AM | No |
| 30 | 1000 | Retail Store | 1:10 PM | No |

**Problem 3: Support Vector Machine (SVM) / Random Forest (RF):**

Problem: Classify emails as spam or not spam based on their content.

| **Email ID** | **Average Word Length** | **Number of Uppercase Characters** | **Number of Exclamation Marks** | **Spam/Not Spam** |
| --- | --- | --- | --- | --- |
| 1 | 5.2 | 3 | 0 | Spam |
| 2 | 4.9 | 1 | 1 | Not Spam |
| 3 | 6.1 | 5 | 2 | Spam |
| 4 | 4.5 | 2 | 0 | Not Spam |
| 5 | 5.8 | 4 | 3 | Spam |
| 6 | 5.0 | 2 | 0 | Not Spam |
| 7 | 5.5 | 3 | 2 | Spam |
| 8 | 5.9 | 1 | 1 | Not Spam |
| 9 | 4.7 | 4 | 4 | Spam |
| 10 | 5.6 | 2 | 0 | Not Spam |
| 11 | 4.8 | 3 | 1 | Spam |
| 12 | 6.0 | 1 | 0 | Not Spam |
| 13 | 5.7 | 4 | 2 | Spam |
| 14 | 4.3 | 2 | 0 | Not Spam |
| 15 | 5.9 | 5 | 3 | Spam |
| 16 | 5.1 | 1 | 1 | Not Spam |
| 17 | 6.2 | 3 | 2 | Spam |
| 18 | 4.6 | 2 | 0 | Not Spam |
| 19 | 5.5 | 4 | 4 | Spam |
| 20 | 5.0 | 1 | 0 | Not Spam |
| 21 | 6.0 | 3 | 3 | Spam |
| 22 | 4.9 | 2 | 1 | Not Spam |
| 23 | 5.8 | 4 | 2 | Spam |
| 24 | 5.2 | 1 | 0 | Not Spam |
| 25 | 6.1 | 5 | 4 | Spam |
| 26 | 4.7 | 2 | 0 | Not Spam |
| 27 | 5.6 | 3 | 2 | Spam |
| 28 | 5.0 | 1 | 1 | Not Spam |
| 29 | 5.4 | 4 | 3 | Spam |
| 30 | 4.8 | 2 | 0 | Not Spam |

**Problem 4: Apply K-Means Clustering Using the following Dataset:**

| **Document ID** | **Feature 1** | **Feature 2** | **Feature 3** | **Feature 4** | **Cluster** |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 1 | 0.1 | A |
| 2 | 2.5 | 4 | 1.5 | 0.2 | A |
| 3 | 2.1 | 3.8 | 1.2 | 0.15 | A |
| 4 | 2.3 | 4.1 | 1.3 | 0.18 | A |
| 5 | 2.4 | 3.9 | 1.4 | 0.12 | A |
| 6 | 5.9 | 8 | 6 | 1.8 | B |
| 7 | 6.2 | 7.8 | 5.8 | 1.9 | B |
| 8 | 6 | 8 | 5.5 | 1.7 | B |
| 9 | 5.8 | 7.6 | 5.7 | 1.5 | B |
| 10 | 6.1 | 8.2 | 5.9 | 2 | B |
| 11 | 9 | 11 | 10 | 3 | C |
| 12 | 9.5 | 10.5 | 9.8 | 2.8 | C |
| 13 | 8.8 | 11.2 | 10.1 | 2.7 | C |
| 14 | 10 | 10 | 9.5 | 2.9 | C |
| 15 | 9.2 | 11.5 | 10.5 | 3.1 | C |
| 16 | 1.5 | 2.9 | 0.8 | 0.09 | A |
| 17 | 1.8 | 2.7 | 1.1 | 0.1 | A |
| 18 | 6.5 | 8.5 | 6.2 | 1.9 | B |
| 19 | 6.3 | 7.9 | 6.1 | 1.8 | B |
| 20 | 9.1 | 11.3 | 9.9 | 2.6 | C |
| 21 | 9.3 | 11.6 | 10.3 | 2.8 | C |
| 22 | 2.2 | 3.7 | 1.1 | 0.13 | A |
| 23 | 6.4 | 7.7 | 5.9 | 1.8 | B |
| 24 | 9.4 | 10.8 | 9.7 | 2.7 | C |
| 25 | 10.1 | 11.2 | 9.8 | 3 | C |
| 26 | 1.9 | 3 | 1.2 | 0.14 | A |
| 27 | 6.7 | 7.5 | 5.6 | 1.7 | B |
| 28 | 9.6 | 10.9 | 9.9 | 2.9 | C |
| 29 | 2.3 | 3.5 | 1.3 | 0.16 | A |
| 30 | 10.2 | 11.5 | 10.2 | 3.1 | C |

**Problem 5: Naive Bayes Example Dataset (Fruit Classification):**

| **Fruit ID** | **Colour** | **Size** | **Shape** | **Texture** | **Is Apple** |
| --- | --- | --- | --- | --- | --- |
| 1 | Red | Small | Round | Smooth | Yes |
| 2 | Red | Small | Round | Bumpy | Yes |
| 3 | Green | Small | Round | Smooth | Yes |
| 4 | Green | Large | Oval | Smooth | No |
| 5 | Yellow | Large | Oval | Bumpy | No |
| 6 | Red | Small | Round | Smooth | Yes |
| 7 | Yellow | Small | Round | Bumpy | Yes |
| 8 | Green | Large | Oval | Smooth | No |
| 9 | Red | Large | Oval | Bumpy | No |
| 10 | Yellow | Small | Round | Smooth | Yes |
| 11 | Green | Small | Round | Smooth | Yes |
| 12 | Yellow | Large | Oval | Bumpy | No |
| 13 | Red | Small | Round | Smooth | Yes |
| 14 | Yellow | Small | Round | Bumpy | Yes |
| 15 | Green | Large | Oval | Smooth | No |
| 16 | Red | Large | Oval | Bumpy | No |
| 17 | Yellow | Small | Round | Smooth | Yes |
| 18 | Green | Small | Round | Smooth | Yes |
| 19 | Yellow | Large | Oval | Bumpy | No |
| 20 | Red | Small | Round | Smooth | Yes |
| 21 | Yellow | Small | Round | Bumpy | Yes |
| 22 | Green | Large | Oval | Smooth | No |
| 23 | Red | Large | Oval | Bumpy | No |
| 24 | Yellow | Small | Round | Smooth | Yes |
| 25 | Green | Small | Round | Smooth | Yes |
| 26 | Yellow | Large | Oval | Bumpy | No |
| 27 | Red | Small | Round | Smooth | Yes |
| 28 | Yellow | Small | Round | Bumpy | Yes |
| 29 | Green | Large | Oval | Smooth | No |
| 30 | Red | Large | Oval | Bumpy | No |

### Practice Problem: Customer Segmentation for an E-commerce Platform

#### Objective:

Build and compare the performance of Random Forest and SVM classifiers to segment customers based on their shopping behavior.

#### Dataset Description:

You have an e-commerce dataset containing information about customers and their purchasing behavior. The goal is to classify customers into three segments: Low Spenders, Medium Spenders, and High Spenders.

#### Features:

1. age: Age of the customer (in years)
2. annual\_income: Annual income of the customer (in thousand dollars)
3. total\_spent: Total amount spent by the customer (in thousand dollars)
4. shopping\_frequency: Number of shopping visits in the last year
5. avg\_transaction\_value: Average value of each transaction (in dollars)

#### Target:

segment: Customer segment, which can be one of the following:

* 0: Low Spenders
* 1: Medium Spenders
* 2: High Spenders

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **age** | **annual\_income** | **total\_spent** | **shopping\_frequency** | **avg\_transaction\_value** | **segment** |
| 22 | 112 | 8.44 | 81 | 0.10 | 0 |
| 49 | 69 | 11.54 | 59 | 0.20 | 0 |
| 38 | 136 | 36.97 | 29 | 1.27 | 1 |
| 56 | 95 | 43.10 | 79 | 0.55 | 1 |
| 59 | 101 | 6.63 | 68 | 0.10 | 0 |
| 34 | 68 | 60.11 | 74 | 0.81 | 2 |
| 58 | 98 | 24.27 | 80 | 0.30 | 1 |
| 31 | 56 | 14.16 | 59 | 0.24 | 0 |
| 43 | 124 | 44.46 | 40 | 1.11 | 1 |
| 44 | 116 | 42.67 | 37 | 1.15 | 1 |
| 28 | 45 | 98.80 | 34 | 2.91 | 2 |
| 61 | 149 | 80.61 | 83 | 0.97 | 2 |
| 66 | 145 | 85.24 | 6 | 14.21 | 2 |
| 66 | 79 | 40.57 | 28 | 1.45 | 1 |
| 68 | 114 | 83.43 | 87 | 0.96 | 2 |
| 68 | 136 | 57.77 | 35 | 1.65 | 2 |
| 45 | 144 | 50.89 | 13 | 3.91 | 2 |
| 55 | 95 | 27.65 | 3 | 9.22 | 1 |
| 29 | 138 | 53.44 | 68 | 0.79 | 2 |
| 37 | 61 | 88.63 | 29 | 3.06 | 2 |
| 33 | 101 | 45.98 | 34 | 1.35 | 1 |
| 24 | 79 | 78.52 | 96 | 0.82 | 2 |
| 52 | 54 | 33.52 | 1 | 33.52 | 1 |
| 38 | 59 | 98.41 | 68 | 1.45 | 2 |
| 32 | 110 | 18.60 | 44 | 0.42 | 0 |
| 18 | 49 | 58.28 | 11 | 5.30 | 2 |
| 18 | 91 | 96.32 | 56 | 1.72 | 2 |
| 62 | 26 | 44.99 | 62 | 0.73 | 1 |
| 46 | 56 | 40.82 | 1 | 40.82 | 1 |
| 23 | 137 | 8.95 | 73 | 0.12 | 0 |
| 65 | 48 | 70.80 | 74 | 0.96 | 2 |
| 52 | 85 | 45.03 | 86 | 0.52 | 1 |
| 66 | 91 | 87.29 | 67 | 1.30 | 2 |
| 49 | 148 | 80.47 | 36 | 2.24 | 2 |
| 65 | 118 | 97.09 | 1 | 97.09 | 2 |
| 51 | 78 | 22.82 | 69 | 0.33 | 1 |
| 53 | 121 | 13.34 | 48 | 0.28 | 0 |
| 33 | 80 | 5.82 | 70 | 0.08 | 0 |
| 40 | 92 | 98.97 | 58 | 1.71 | 2 |
| 49 | 51 | 94.18 | 70 | 1.34 | 2 |
| 55 | 72 | 30.94 | 27 | 1.15 | 1 |
| 20 | 107 | 77.16 | 2 | 38.58 | 2 |
| 65 | 83 | 24.46 | 10 | 2.45 | 1 |
| 34 | 97 | 23.56 | 99 | 0.24 | 1 |
| 43 | 132 | 11.50 | 28 | 0.41 | 0 |
| 27 | 27 | 96.12 | 91 | 1.06 | 2 |