Case-Study_2

November 24, 2024

1 Data Science

1.1 Supervised Learning I

1.1.1 Case Study 2

1.1.2 Objectives:

- Learn to handle missing values
- Learn to fit a DT decision tree and compare its accuracy with a RF random forest classifier.

[71]:

```
import pandas as pd
import numpy as np

data = pd.read_csv('horse.csv')
print("Data information: ")
data.info()

# Check Missing Values
# Count missing values in each column
missing_values = data.isnull().sum()
print("Missing Values:\n", missing_values)

print("\nPrint the first 5 rows:")
print(data.head())
```

```
Data information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 28 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   surgery                299 non-null    object
 1   age                    299 non-null    object
 2   hospital_number        299 non-null    int64
 3   rectal_temp            239 non-null    float64
 4   pulse                  275 non-null    float64
 5   respiratory_rate       241 non-null    float64
 6   temp_of_extremities    243 non-null    object
 7   peripheral_pulse       230 non-null    object
 8   mucous_membrane        252 non-null    object
 9   capillary_refill_time  267 non-null    object
 10  pain                   244 non-null    object
 11  peristalsis            255 non-null    object
 12  abdominal_distention   243 non-null    object
 13  nasogastric_tube       195 non-null    object
 14  nasogastric_reflux     193 non-null    object
 15  nasogastric_reflux_ph  53 non-null     float64
 16  rectal_exam_feces      197 non-null    object
 17  abdomen                181 non-null    object
 18  packed_cell_volume     270 non-null    float64
 19  total_protein          266 non-null    float64
 20  abdomo_appearance      134 non-null    object
 21  abdomo_protein         101 non-null    float64
 22  outcome                299 non-null    object
 23  surgical_lesion        299 non-null    object
 24  lesion_1               299 non-null    int64
 25  lesion_2               299 non-null    int64
 26  lesion_3               299 non-null    int64
 27  cp_data                299 non-null    object
dtypes: float64(7), int64(4), object(17)
memory usage: 65.5+ KB
Missing Values:
 surgery                    0
age                        0
hospital_number            0
rectal_temp               60
pulse                     24
respiratory_rate          58
temp_of_extremities       56
peripheral_pulse          69
mucous_membrane           47
capillary_refill_time     32
pain                      55
peristalsis               44
abdominal_distention      56
nasogastric_tube         104
nasogastric_reflux       106
nasogastric_reflux_ph    246
rectal_exam_feces        102
abdomen                  118
packed_cell_volume        29
total_protein             33
abdomo_appearance        165
abdomo_protein           198
outcome                    0
surgical_lesion            0
lesion_1                   0
```

```
lesion_2                   0
lesion_3                   0
cp_data                    0
dtype: int64

Print the first 5 rows:
  surgery    age  hospital_number  rectal_temp  pulse  respiratory_rate  \
0      no  adult           530101         38.5   66.0              28.0
1     yes  adult           534817         39.2   88.0              20.0
2      no  adult           530334         38.3   40.0              24.0
3     yes  young          5290409         39.1  164.0              84.0
4      no  adult           530255         37.3  104.0              35.0

  temp_of_extremities peripheral_pulse mucous_membrane capillary_refill_time  \
0                cool          reduced             NaN            more_3_sec
1                 NaN              NaN   pale_cyanotic            less_3_sec
2              normal           normal       pale_pink            less_3_sec
3                cold           normal   dark_cyanotic            more_3_sec
4                 NaN              NaN   dark_cyanotic            more_3_sec

   … packed_cell_volume total_protein abdomo_appearance abdomo_protein  \
0  …               45.0           8.4               NaN            NaN
1  …               50.0          85.0            cloudy            2.0
2  …               33.0           6.7               NaN            NaN
3  …               48.0           7.2     serosanguious            5.3
4  …               74.0           7.4               NaN            NaN

      outcome  surgical_lesion lesion_1 lesion_2  lesion_3  cp_data
0        died               no    11300        0         0       no
1  euthanized               no     2208        0         0       no
2       lived               no        0        0         0      yes
3        died              yes     2208        0         0      yes
4        died               no     4300        0         0       no

[5 rows x 28 columns]
```

1.1.3 Understanding the Data

The dataset contains 299 samples (rows) and 28 columns (features), which include a mix of: * Categorical variables (e.g., surgery, outcome, age). * Numerical variables (e.g., rectal_temp, pulse). * Missing values across many columns.

The goal is to predict the survival of a horse (outcome) based on medical observations.

1.1.4 Describing the Columns

Target Variable

- outcome: The survival status of the horse. It has three categories: lived, died, and euthanized.

#### Other Features

- Categorical (17 columns):
  - Examples: surgery, age, temp_of_extremities, peripheral_pulse, surgical_lesion.
- Numerical (7 columns):
  - Examples: rectal_temp, pulse, packed_cell_volume.
- Other Metadata Columns (4 columns):
  - Examples: hospital_number (likely a unique identifier), lesion_1, lesion_2, lesion_3.

1.1.5 Initial Observations

- Several columns have missing values, such as:
  - rectal_temp (239 non-null out of 299).
  - nasogastric_reflux_ph (53 non-null, very sparse).
- Some features like hospital_number appear irrelevant for predicting survival.
- Categorical data needs encoding for use in machine learning models.

[72]:

```
# Replace Missing Values

from sklearn.impute import SimpleImputer

# Separate categorical and numerical columns
categorical_cols = data.select_dtypes(include=['object']).columns
numerical_cols = data.select_dtypes(include=['float64', 'int64']).columns

# Replace missing values in categorical columns with the most frequent value
cat_imputer = SimpleImputer(strategy='most_frequent')
data[categorical_cols] = cat_imputer.fit_transform(data[categorical_cols])

# Replace missing values in numerical columns with the most frequent value
num_imputer = SimpleImputer(strategy='most_frequent')
data[numerical_cols] = num_imputer.fit_transform(data[numerical_cols])



data.info()
data.head(50)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 28 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   surgery                299 non-null    object
 1   age                    299 non-null    object
 2   hospital_number        299 non-null    float64
 3   rectal_temp            299 non-null    float64
 4   pulse                  299 non-null    float64
 5   respiratory_rate       299 non-null    float64
 6   temp_of_extremities    299 non-null    object
 7   peripheral_pulse       299 non-null    object 
 8   mucous_membrane        299 non-null    object 
 9   capillary_refill_time  299 non-null    object 
 10  pain                   299 non-null    object 
 11  peristalsis            299 non-null    object 
 12  abdominal_distention   299 non-null    object 
 13  nasogastric_tube       299 non-null    object 
 14  nasogastric_reflux     299 non-null    object 
 15  nasogastric_reflux_ph  299 non-null    float64
 16  rectal_exam_feces      299 non-null    object 
 17  abdomen                299 non-null    object 
 18  packed_cell_volume     299 non-null    float64
 19  total_protein          299 non-null    float64
 20  abdomo_appearance      299 non-null    object 
 21  abdomo_protein         299 non-null    float64
 22  outcome                299 non-null    object 
 23  surgical_lesion        299 non-null    object 
 24  lesion_1               299 non-null    float64
 25  lesion_2               299 non-null    float64
 26  lesion_3               299 non-null    float64
 27  cp_data                299 non-null    object 
dtypes: float64(11), object(17)
memory usage: 65.5+ KB
```

```
[72]:    surgery    age  hospital_number  rectal_temp  pulse  respiratory_rate  \
      0        no  adult         530101.0         38.5   66.0              28.0   
      1       yes  adult         534817.0         39.2   88.0              20.0   
      2        no  adult         530334.0         38.3   40.0              24.0   
      3       yes  young        5290409.0         39.1  164.0              84.0   
      4        no  adult         530255.0         37.3  104.0              35.0   
      5        no  adult         528355.0         38.0   48.0              20.0   
      6       yes  adult         526802.0         37.9   48.0              16.0   
      7       yes  adult         529607.0         38.0   60.0              20.0   
      8        no  adult         530051.0         38.0   80.0              36.0   
      9        no  young        5299629.0         38.3   90.0              20.0   
      10      yes  adult         528548.0         38.1   66.0              12.0   
      11       no  adult         527927.0         39.1   72.0              52.0   
      12      yes  adult         528031.0         37.2   42.0              12.0   
      13       no  young        5291329.0         38.0   92.0              28.0   
      14      yes  adult         534917.0         38.2   76.0              28.0   
      15      yes  adult         530233.0         37.6   96.0              48.0   
      16      yes  young        5301219.0         38.0  128.0              36.0   
      17       no  adult         526639.0         37.5   48.0              24.0   
      18      yes  adult        5290481.0         37.6   64.0              21.0   
      19       no  adult         532110.0         39.4  110.0              35.0   
      20      yes  adult         530157.0         39.9   72.0              60.0   
      21       no  adult         529340.0         38.4   48.0              16.0   
```

```
22     yes  adult         521681.0         38.6   42.0              34.0
23     yes  young         534998.0         38.3  130.0              60.0
24     yes  adult         533692.0         38.1   60.0              12.0
25      no  adult         529518.0         37.8   60.0              42.0
26     yes  adult         530526.0         38.3   72.0              30.0
27     yes  adult         528653.0         37.8   48.0              12.0
28     yes  adult        5279442.0         38.0   48.0              20.0
29      no  adult         535415.0         37.7   48.0              20.0
30      no  adult         529475.0         37.7   96.0              30.0
31      no  adult         530242.0         37.2  108.0              12.0
32     yes  adult         529427.0         37.2   60.0              20.0
33     yes  adult         529663.0         38.2   64.0              28.0
34     yes  adult         529796.0         38.0  100.0              30.0
35      no  adult         528812.0         38.0  104.0              24.0
36      no  adult         529493.0         38.3  112.0              16.0
37     yes  adult         533847.0         37.8   72.0              20.0
38      no  adult         528996.0         38.6   52.0              20.0
39     yes  young        5277409.0         39.2  146.0              96.0
40     yes  adult         529498.0         38.0   88.0              20.0
41      no  young        5288249.0         39.0  150.0              72.0
42      no  adult         530301.0         38.0   60.0              12.0
43     yes  adult         534069.0         38.0  120.0              20.0
44     yes  adult         535407.0         35.4  140.0              24.0
45      no  adult         529827.0         38.0  120.0              20.0
46     yes  adult         529888.0         37.9   60.0              15.0
47      no  adult         529821.0         37.5   48.0              16.0
48     yes  adult         528890.0         38.9   80.0              44.0
49      no  adult         529642.0         37.2   84.0              48.0

   temp_of_extremities peripheral_pulse mucous_membrane capillary_refill_time  \
0                 cool          reduced     normal_pink            more_3_sec
1                 cool           normal   pale_cyanotic            less_3_sec
2               normal           normal       pale_pink            less_3_sec
3                 cold           normal   dark_cyanotic            more_3_sec
4                 cool           normal   dark_cyanotic            more_3_sec
5                 warm           normal       pale_pink            less_3_sec
6               normal           normal     normal_pink            less_3_sec
7                 cool           normal     normal_pink            less_3_sec
8                 cool           absent       pale_pink            less_3_sec
9               normal           normal     normal_pink            less_3_sec
10                cool          reduced      bright_red            less_3_sec
11                warm           normal     bright_pink            less_3_sec
12                warm           normal     normal_pink            less_3_sec
13              normal           normal     bright_pink            less_3_sec
14                cool           normal     normal_pink            less_3_sec
15                cool           normal   pale_cyanotic            less_3_sec
16                cool          reduced   pale_cyanotic            more_3_sec
```

```
17                cool           normal     normal_pink            less_3_sec
18              normal           normal     bright_pink            less_3_sec
19                cold          reduced   dark_cyanotic            less_3_sec
20              normal           normal      bright_red            more_3_sec
21              normal           normal     normal_pink            less_3_sec
22                warm           normal   pale_cyanotic            less_3_sec
23                cool          reduced     normal_pink            less_3_sec
24                cool          reduced       pale_pink            less_3_sec
25                cool           normal     normal_pink            less_3_sec
26                cold          reduced       pale_pink            more_3_sec
27                cool           normal     normal_pink            less_3_sec
28                cool           normal     normal_pink            less_3_sec
29                warm           normal     normal_pink            less_3_sec
30                cool          reduced   pale_cyanotic            more_3_sec
31                cool          reduced   pale_cyanotic            more_3_sec
32                warm           normal     normal_pink            less_3_sec
33              normal           normal     normal_pink            less_3_sec
34                cool          reduced   pale_cyanotic            more_3_sec
35                cold          reduced       pale_pink            more_3_sec
36                cool          reduced      bright_red            more_3_sec
37                cool          reduced     normal_pink            less_3_sec
38              normal           normal     normal_pink            less_3_sec
39                cool           normal     normal_pink            less_3_sec
40                cool          reduced   dark_cyanotic            more_3_sec
41                cool           normal     normal_pink            less_3_sec
42                cool           normal       pale_pink            less_3_sec
43                cool           absent   pale_cyanotic            less_3_sec
44                cool          reduced   pale_cyanotic            more_3_sec
45                cold          reduced   pale_cyanotic            more_3_sec
46                cool           normal   pale_cyanotic            more_3_sec
47              normal           normal     normal_pink            less_3_sec
48                cool          reduced       pale_pink            more_3_sec
49                cool          reduced      bright_red            more_3_sec

    … packed_cell_volume total_protein abdomo_appearance abdomo_protein  \
0   …               45.0           8.4            cloudy            2.0
1   …               50.0          85.0            cloudy            2.0
2   …               33.0           6.7            cloudy            2.0
3   …               48.0           7.2     serosanguious            5.3
4   …               74.0           7.4            cloudy            2.0
5   …               37.0           6.5            cloudy            2.0
6   …               37.0           7.0            cloudy            2.0
7   …               44.0           8.3            cloudy            2.0
8   …               38.0           6.2            cloudy            2.0
9   …               40.0           6.2             clear            2.2
10  …               44.0           6.0            cloudy            3.6
11  …               50.0           7.8            cloudy            2.0
```

```
12  …               37.0           7.0            cloudy            2.0
13  …               37.0           6.1             clear            2.0
14  …               46.0          81.0             clear            2.0
15  …               45.0           6.8            cloudy            2.0
16  …               53.0           7.8     serosanguious            4.7
17  …               37.0           6.5            cloudy            2.0
18  …               40.0           7.0             clear            2.0
19  …               55.0           8.7            cloudy            2.0
20  …               46.0           6.1            cloudy            2.0
21  …               49.0           6.8            cloudy            2.0
22  …               48.0           7.2            cloudy            2.0
23  …               50.0          70.0            cloudy            2.0
24  …               51.0          65.0            cloudy            2.0
25  …               37.0           6.5            cloudy            2.0
26  …               43.0           7.0            cloudy            3.9
27  …               37.0           5.5            cloudy            1.3
28  …               37.0           6.5            cloudy            2.0
29  …               45.0          76.0            cloudy            2.0
30  …               66.0           7.5            cloudy            2.0
31  …               52.0           8.2     serosanguious            7.4
32  …               43.0           6.6            cloudy            2.0
33  …               49.0           8.6            cloudy            6.6
34  …               52.0           6.6            cloudy            2.0
35  …               73.0           8.4            cloudy            2.0
36  …               51.0           6.0            cloudy            1.0
37  …               56.0          80.0             clear            2.0
38  …               32.0           6.6             clear            5.0
39  …               37.0           6.5            cloudy            2.0
40  …               63.0           6.5     serosanguious            2.0
41  …               47.0           8.5            cloudy            0.1
42  …               47.0           7.0            cloudy            2.0
43  …               52.0          67.0            cloudy            2.0
44  …               57.0          69.0     serosanguious            2.0
45  …               60.0           6.5     serosanguious            2.0
46  …               65.0           7.5            cloudy            2.0
47  …               37.0           6.5            cloudy            2.0
48  …               54.0           6.5     serosanguious            2.0
49  …               73.0           5.5            cloudy            4.1

       outcome  surgical_lesion lesion_1 lesion_2  lesion_3  cp_data
0         died               no  11300.0      0.0       0.0       no
1   euthanized               no   2208.0      0.0       0.0       no
2        lived               no      0.0      0.0       0.0      yes
3         died              yes   2208.0      0.0       0.0      yes
4         died               no   4300.0      0.0       0.0       no
5        lived               no      0.0      0.0       0.0       no
6        lived              yes   3124.0      0.0       0.0       no
```

```
7         died              yes   2208.0       0.0       0.0       no
8   euthanized              yes   3205.0       0.0       0.0       no
9        lived               no      0.0       0.0       0.0      yes
10       lived              yes   2124.0       0.0       0.0      yes
11       lived              yes   2111.0       0.0       0.0       no
12       lived               no   4124.0       0.0       0.0       no
13        died               no      0.0       0.0       0.0      yes
14       lived              yes   2112.0       0.0       0.0       no
15        died              yes   3207.0       0.0       0.0       no
16        died               no   1400.0       0.0       0.0      yes
17       lived               no      0.0       0.0       0.0       no
18       lived              yes   4205.0       0.0       0.0      yes
19       lived               no      0.0       0.0       0.0       no
20       lived              yes   2111.0       0.0       0.0       no
21       lived               no      0.0       0.0       0.0       no
22       lived              yes   3111.0       0.0       0.0       no
23       lived              yes   3111.0       0.0       0.0       no
24       lived              yes   3111.0       0.0       0.0       no
25       lived               no      0.0       0.0       0.0       no
26       lived              yes   3111.0       0.0       0.0      yes
27       lived               no   4122.0       0.0       0.0      yes
28        died               no   4300.0       0.0       0.0       no
29       lived               no      0.0       0.0       0.0       no
30        died              yes   4205.0       0.0       0.0       no
31  euthanized              yes   2207.0       0.0       0.0      yes
32       lived              yes   2209.0       0.0       0.0       no
33       lived              yes   2208.0       0.0       0.0      yes
34       lived              yes   1124.0       0.0       0.0       no
35  euthanized              yes   7111.0       0.0       0.0       no
36  euthanized               no   5205.0       0.0       0.0      yes
37       lived              yes   3111.0       0.0       0.0       no
38       lived               no      0.0       0.0       0.0      yes
39        died              yes   2113.0       0.0       0.0       no
40        died              yes   4205.0       0.0       0.0       no
41       lived              yes   9400.0       0.0       0.0      yes
42       lived               no   3111.0       0.0       0.0       no
43  euthanized              yes   3205.0       0.0       0.0       no
44  euthanized              yes   3205.0       0.0       0.0       no
45        died              yes   3205.0       0.0       0.0       no
46       lived              yes   2305.0       0.0       0.0      yes
47       lived               no      0.0       0.0       0.0       no
48        died              yes   7111.0       0.0       0.0       no
49        died               no   4300.0       0.0       0.0      yes

[50 rows x 28 columns]
```

[73]:

```
# Encode Categorical Features

from sklearn.preprocessing import LabelEncoder

# Apply label encoding to categorical columns
label_encoder = LabelEncoder()

for col in categorical_cols:
    data[col] = label_encoder.fit_transform(data[col].astype(str))

data.info()
data.head()
data.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 28 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   surgery                299 non-null    int64
 1   age                    299 non-null    int64
 2   hospital_number        299 non-null    float64
 3   rectal_temp            299 non-null    float64
 4   pulse                  299 non-null    float64
 5   respiratory_rate       299 non-null    float64
 6   temp_of_extremities    299 non-null    int64
 7   peripheral_pulse       299 non-null    int64
 8   mucous_membrane        299 non-null    int64
 9   capillary_refill_time  299 non-null    int64
 10  pain                   299 non-null    int64
 11  peristalsis            299 non-null    int64
 12  abdominal_distention   299 non-null    int64
 13  nasogastric_tube       299 non-null    int64
 14  nasogastric_reflux     299 non-null    int64
 15  nasogastric_reflux_ph  299 non-null    float64
 16  rectal_exam_feces      299 non-null    int64
 17  abdomen                299 non-null    int64
 18  packed_cell_volume     299 non-null    float64
 19  total_protein          299 non-null    float64
 20  abdomo_appearance      299 non-null    int64
 21  abdomo_protein         299 non-null    float64
 22  outcome                299 non-null    int64
 23  surgical_lesion        299 non-null    int64
 24  lesion_1               299 non-null    float64
 25  lesion_2               299 non-null    float64
 26  lesion_3               299 non-null    float64
 27  cp_data                299 non-null    int64
dtypes: float64(11), int64(17)
```

```
memory usage: 65.5 KB
```

[73]:

```
surgery                  0
age                      0
hospital_number          0
rectal_temp              0
pulse                    0
respiratory_rate         0
temp_of_extremities      0
peripheral_pulse         0
mucous_membrane          0
capillary_refill_time    0
pain                     0
peristalsis              0
abdominal_distention     0
nasogastric_tube         0
nasogastric_reflux       0
nasogastric_reflux_ph    0
rectal_exam_feces        0
abdomen                  0
packed_cell_volume       0
total_protein            0
abdomo_appearance        0
abdomo_protein           0
outcome                  0
surgical_lesion          0
lesion_1                 0
lesion_2                 0
lesion_3                 0
cp_data                  0
dtype: int64
```

[74]:

```
# Proceed with Train-Test Split and Modeling

from sklearn.model_selection import train_test_split

# Define predictors (X) and target (y)
X = data.drop(columns=['outcome'])
y = data['outcome']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
 ↪random_state=42)
```

[75]:

```
# Train and evaluate models (Decision Tree and Random Forest)

from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
```

```

# Train a Decision Tree model
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Evaluate the model
y_pred_dt = dt_model.predict(X_test)
dt_accuracy = accuracy_score(y_test, y_pred_dt)
print(f"Decision Tree Accuracy: {dt_accuracy:.2f}")
```

```
Decision Tree Accuracy: 0.67
```

[76]:

```
# Fit Random Forest Classifier

from sklearn.ensemble import RandomForestClassifier

# Train a Random Forest model
rf_model = RandomForestClassifier(random_state=42, n_estimators=100)
rf_model.fit(X_train, y_train)

# Evaluate the model
y_pred_rf = rf_model.predict(X_test)
rf_accuracy = accuracy_score(y_test, y_pred_rf)
print(f"Random Forest Accuracy: {rf_accuracy:.2f}")
```

```
Random Forest Accuracy: 0.78
```

[77]:

```
# Using GridSearch to Optimize the Number of Trees
# For a Random Forest, the number of trees (n_estimators) can significantly␣
 ↪affect performance.
# We can use GridSearchCV to find the optimal value.

from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV

# Define the parameter grid
param_grid = {
    'n_estimators': [10, 25, 50, 70, 75, 80, 100, 125, 150, 200],  # Range of␣
 ↪trees to test
    'max_depth': [None, 5, 8, 10, 12, 20, 30],        # Optional: Test␣
 ↪different tree depths
    'random_state': [42]                    # Ensure reproducibility
}

# Initialize the Random Forest Classifier
rf = RandomForestClassifier()

# Use GridSearchCV to find the best parameters
```

```python
grid_search = GridSearchCV(estimator=rf, param_grid=param_grid, cv=5, scoring='accuracy', verbose=1, n_jobs=-1)

# Fit GridSearchCV
grid_search.fit(X_train, y_train)

# Get the best parameters and model
best_params = grid_search.best_params_
best_rf_model = grid_search.best_estimator_

print("Best Parameters:", best_params)
```

```
Fitting 5 folds for each of 70 candidates, totalling 350 fits
Best Parameters: {'max_depth': 8, 'n_estimators': 200, 'random_state': 42}
```

[78]:

```python
# Evaluate the best model on the test set
best_rf_accuracy = best_rf_model.score(X_test, y_test)
print(f"Random Forest Accuracy (Best Model): {best_rf_accuracy:.2f}")
```

```
Random Forest Accuracy (Best Model): 0.74
```

1.1.6 Objectives Reached

- Handling Missing Values: Successfully replaced missing values with the most frequent value.
- Encoding: Converted all categorical features into numerical form.
- Model Accuracy:
  - Decision Tree: Achieved an accuracy of 0.67, highlighting potential overfitting.
  - Random Forest: Tuned the number of trees to 200, resulting in an improved accuracy of 0.74 due to ensemble learning.

[ ]:

```python
```