Maternal Health Risk Classification:

Leveraging Machine Learning for Enhanced Risk Management

Background

Maternal Mortality Rates

According to the World Health Organization (WHO), approximately 810 women die from preventable causes related to pregnancy and childbirth every day globally.

Complication Rates during Pregnancy and Childbirth

According to the Centers for Disease Control and Prevention (CDC), approximately 700 women die each year in the United States due to pregnancy-related complications, with significant racial and ethnic disparities in maternal mortality rates.

Impact of Risk Stratification on Maternal Outcomes

Implementing risk-based care pathways tailored to the specific needs of high-risk patients has been associated with reductions in maternal morbidity and mortality rates.

Maternal Health Risk Data Set.csv (30.29 kB)

Detail Compact Column 7 of 7 columns

About this file

Add Suggestion

Data has been collected from different hospitals, community clinics, maternal health cares through the IoT based risk monitoring system.

# Age Age	# SystolicBP Upper value of Blood Pressure in mmHg	# DiastolicBP Lower value of Blood Pressure in mmHg	# BS Blood glucose levels is in terms of a molar concentration, mmol/L	# BodyTemp F	# HeartRate A normal resting heart rate in beats per minute.
 10 70	 70 160	 49 100	 6 19	 98 103	 7 90
25	130	80	15	98	86
35	140	90	13	98	70
29	90	70	8	100	80
30	140	85	7	98	70
35	120	60	6.1	98	76
23	140	80	7.01	98	70
23	130	70	7.01	98	78
35	85	60	11	102	86
32	120	90	6.9	98	70
42	130	80	18	98	70

kaggle

Context

“Data has been collected from different hospitals, community clinics, maternal health cares through the IoT based risk monitoring system.

Age: Age in years when a woman is pregnant.

SystolicBP: Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy.

DiastolicBP: Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.

BS: Blood glucose levels is in terms of a molar concentration, mmol/L.

Heart Rate: A normal resting heart rate in beats per minute.

Risk Level: Predicted Risk Intensity Level during pregnancy considering the previous attribute.”

- Kaggle Dataset Description

Cleaning and Preprocessing the Data:

```
[1]  1 import pandas as pd
```

```
[2]  1 github_url = 'https://raw.githubusercontent.com/marymorkos/maternalmortalityhealthriskdata/main/Maternal%20Health%20Risk%20Data%20Set%204.csv'
     2 maternal_df = pd.read_csv(github_url)
```

```
[3]  1 print(maternal_df.head())
```

```
   Age  SystolicBP  DiastolicBP    BS  BodyTemp  HeartRate  RiskLevel
0   25         130           80  15.0      98.0         86  high risk
1   35         140           90  13.0      98.0         70  high risk
2   29          90           70   8.0     100.0         80  high risk
3   30         140           85   7.0      98.0         70  high risk
4   35         120           60   6.1      98.0         76   low risk
```

```
     1 print(maternal_df.dtypes)
```

```
Age              int64
SystolicBP       int64
DiastolicBP      int64
BS             float64
BodyTemp       float64
HeartRate        int64
RiskLevel       object
dtype: object
```

```
[5]  1 print(maternal_df.isnull().sum())
```

```
Age           0
SystolicBP    0
DiastolicBP   0
BS            0
BodyTemp      0
HeartRate     0
RiskLevel     0
```

```
[5] HeartRate      0
    RiskLevel      0
    dtype: int64
```

```
[6]  1 print(maternal_df.duplicated().sum())
```

```
562
```

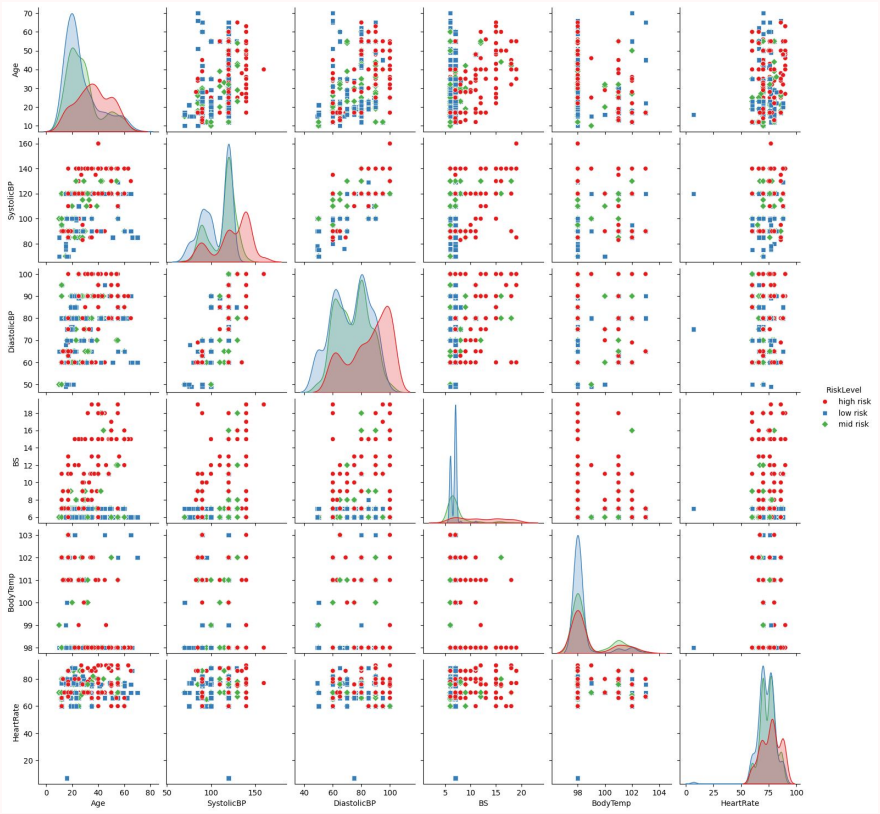
```
[7]  1 print(maternal_df.describe())
```

```
                Age   SystolicBP  DiastolicBP           BS     BodyTemp  \
count  1014.000000  1014.000000  1014.000000  1014.000000  1014.000000
mean     29.871795   113.198225    76.460552     8.725986    98.665089
std      13.474386    18.403913    13.885796     3.293532     1.371384
min      10.000000    70.000000    49.000000     6.000000    98.000000
25%      19.000000   100.000000    65.000000     6.900000    98.000000
50%      26.000000   120.000000    80.000000     7.500000    98.000000
75%      39.000000   120.000000    90.000000     8.000000    98.000000
max      70.000000   160.000000   100.000000    19.000000   103.000000

         HeartRate
count  1014.000000
mean     74.301775
std       8.088702
min       7.000000
25%      70.000000
50%      76.000000
75%      80.000000
max      90.000000
```

More cell actions

```
[8]  1 float_columns = ['BS', 'BodyTemp']
     2 maternal_df[float_columns] = maternal_df[float_columns].astype(int)
     3
     4 maternal_df['RiskLevel'].replace({'low': 1, 'mid': 2, 'high': 3}, inplace=True)
```

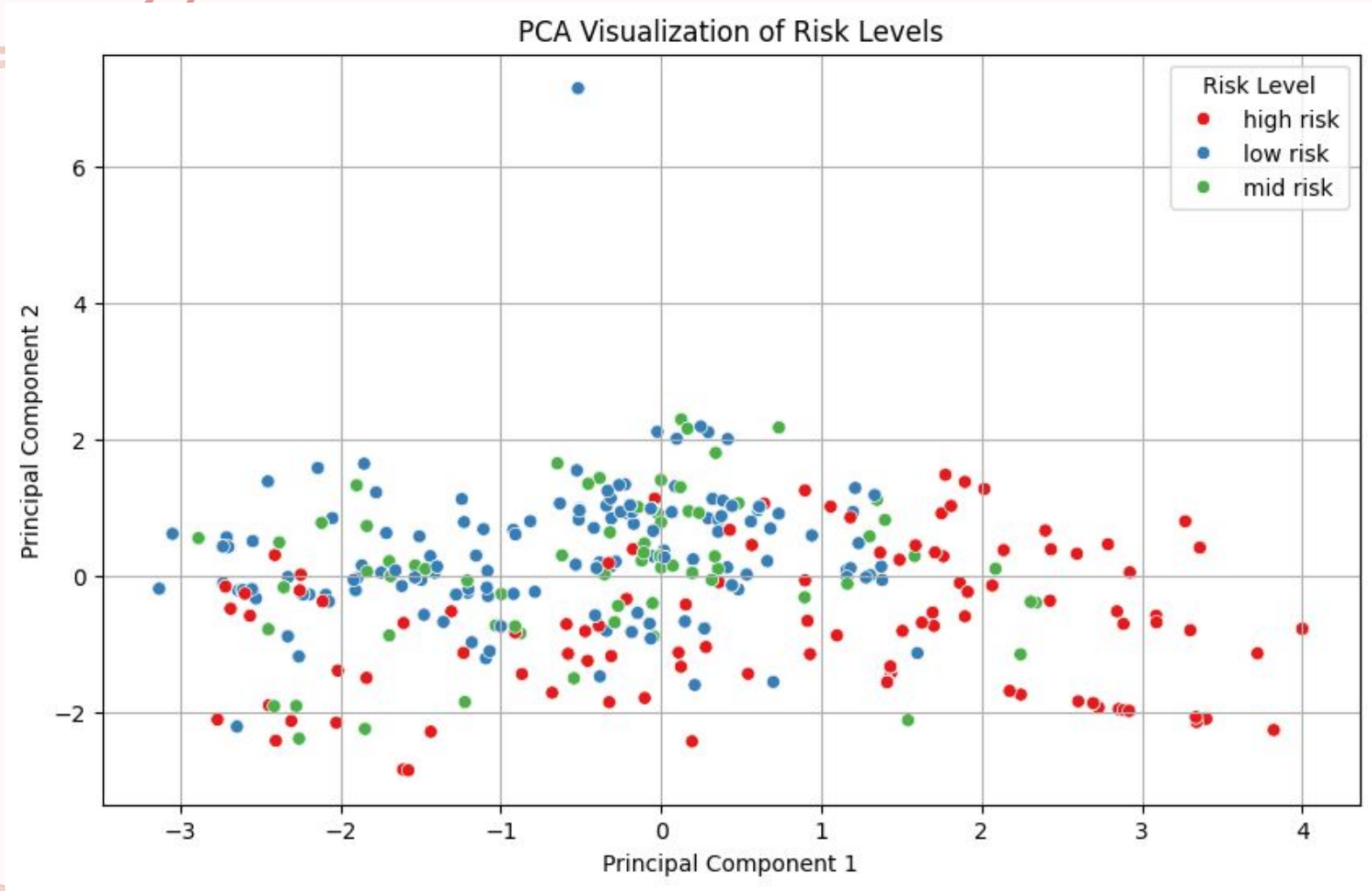
```
1 sns.pairplot(maternal_df, hue='RiskLevel', markers=["o", "s", "D"], palette="Set1")
2 plt.show()
```



Accuracy: 0.6157635467980296

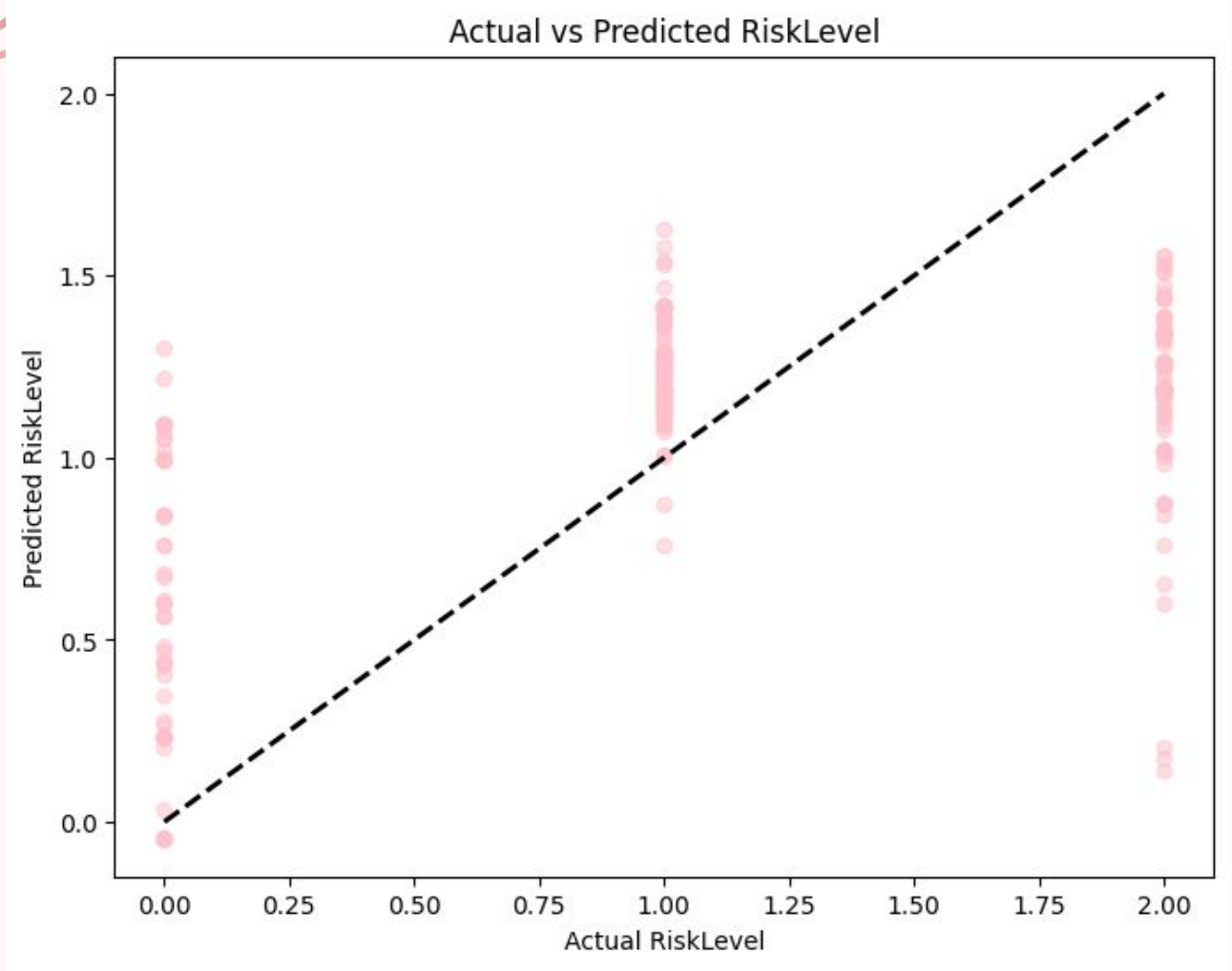
PCA MODEL

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

# Separate features and target variable
X = maternal_df.drop(columns=['RiskLevel'])
y = maternal_df['RiskLevel']

# Standardize the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Perform PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

# Create a DataFrame for the PCA results
pca_df = pd.DataFrame(data=X_pca, columns=['PC1', 'PC2'])
pca_df['RiskLevel'] = y

# Plot PCA results
plt.figure(figsize=(10, 6))
sns.scatterplot(data=pca_df, x='PC1', y='PC2', hue='RiskLevel', palette='Set1', markers=["o", "s", "D"])
plt.title('PCA Visualization of Risk Levels')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.legend(title='Risk Level')
plt.grid(True)
plt.show()

accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```


PCA Visualization of Risk Levels
Risk Level
high risk
low risk
mid risk
Principal Component 2
Principal Component 1
6
4
2
0
−2
−3
−2
−1
0
1
2
3
4

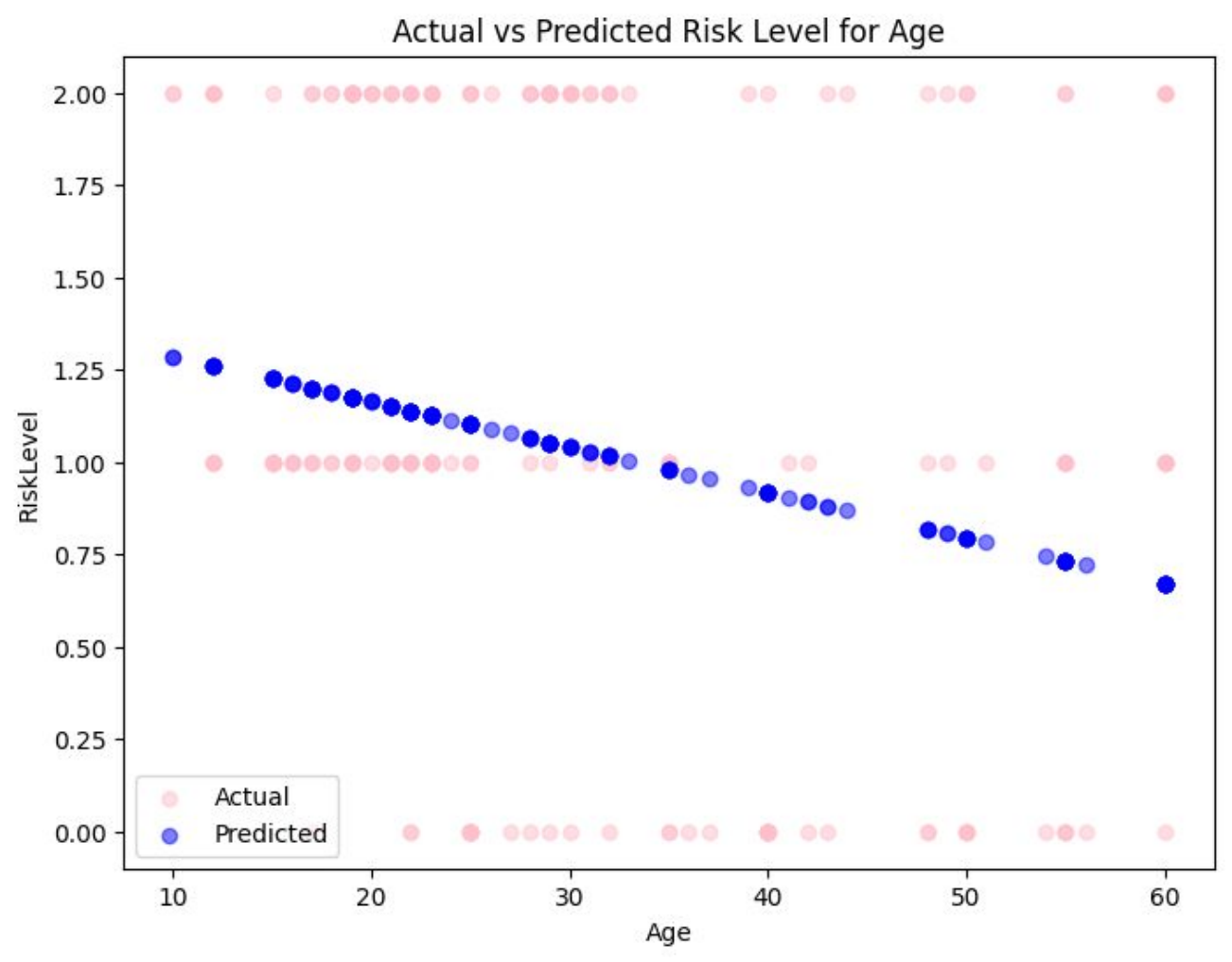
Root Mean Squared Error (RMSE): 0.6595912891051892

LINEAR REGRESSION MODEL for Risk Level


Actual vs Predicted RiskLevel
Predicted RiskLevel
Actual RiskLevel
0.0
0.5
1.0
1.5
2.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

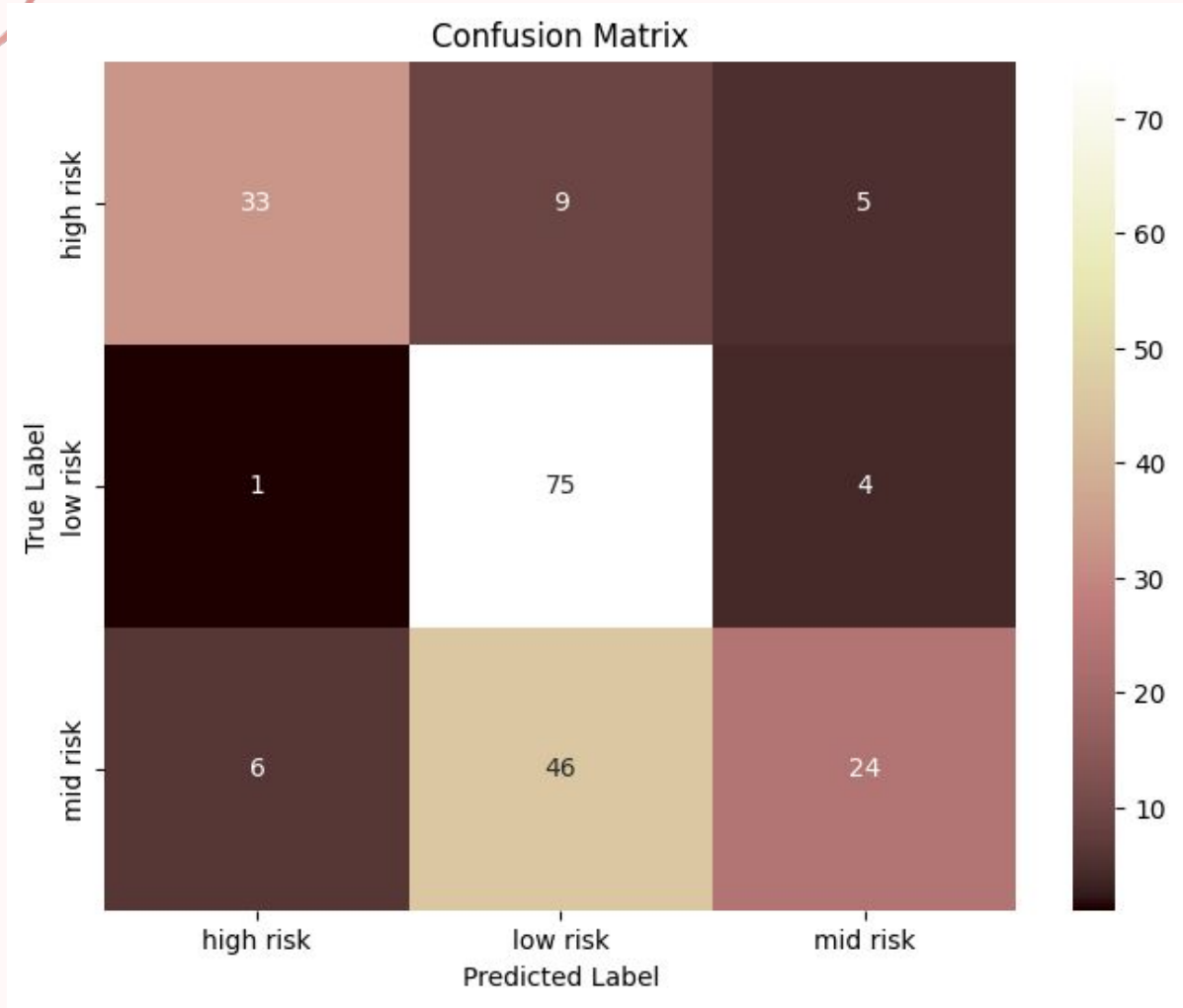
Root Mean Squared Error (RMSE): 0.755122134918198

LINEAR REGRESSION MODEL for Risk Level vs. Age


Actual vs Predicted Risk Level for Age
RiskLevel
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Actual
Predicted
10
20
30
40
50
60
Age

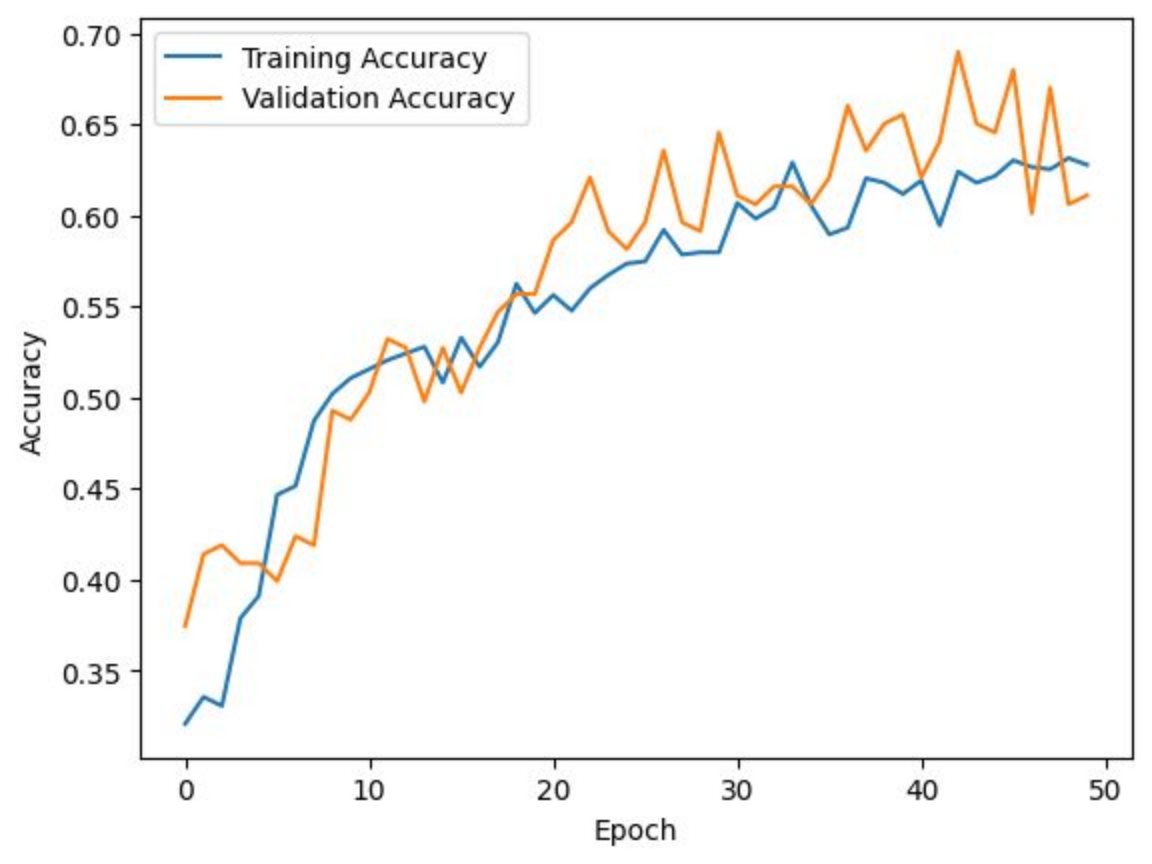
Accuracy: 0.6502463054187192

Logistic Regression


Confusion Matrix
high risk
low risk
mid risk
True Label
33
9
5
1
75
4
6
46
24
high risk
low risk
mid risk
Predicted Label
70
60
50
40
30
20
10

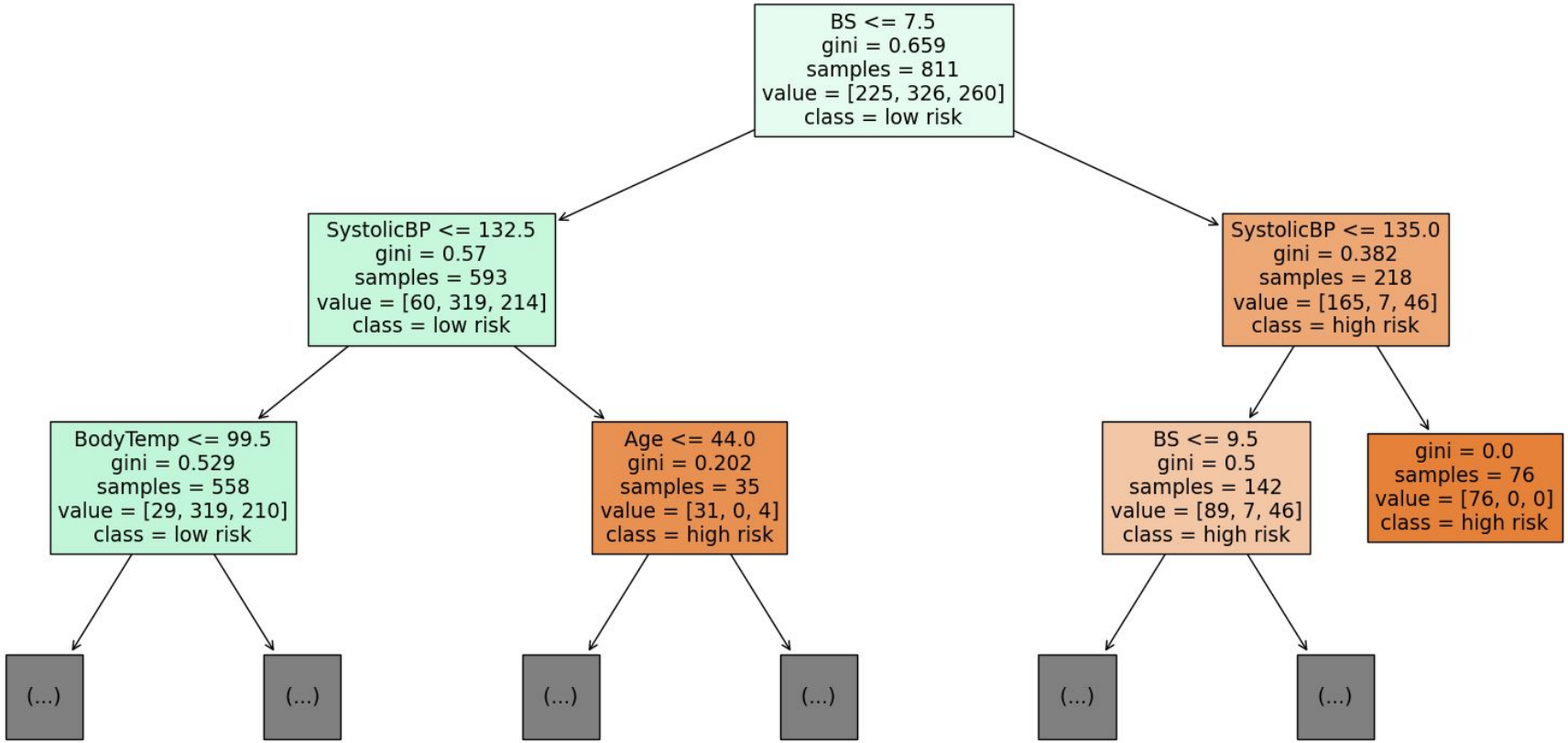
7/7 [==============================] - 0s 3ms/step - loss: 0.8183 - accuracy: 0.6108
Accuracy: 0.6108374595642090

Neural Network


Training Accuracy
Validation Accuracy
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
Accuracy
0
10
20
30
40
50
Epoch

Accuracy: 0.7783251231527094

Classification Tree


BS <= 7.5
gini = 0.659
samples = 811
value = [225, 326, 260]
class = low risk
SystolicBP <= 132.5
gini = 0.57
samples = 593
value = [60, 319, 214]
class = low risk
SystolicBP <= 135.0
gini = 0.382
samples = 218
value = [165, 7, 46]
class = high risk
BodyTemp <= 99.5
gini = 0.529
samples = 558
value = [29, 319, 210]
class = low risk
Age <= 44.0
gini = 0.202
samples = 35
value = [31, 0, 4]
class = high risk
BS <= 9.5
gini = 0.5
samples = 142
value = [89, 7, 46]
class = high risk
gini = 0.0
samples = 76
value = [76, 0, 0]
class = high risk
(...)
(...)
(...)
(...)
(...)
(...)

0.7783

Classification Tree

0.755

LINEAR REGRESSION MODEL for Risk Level vs. Age

0.659

LINEAR REGRESSION MODEL for Risk Level

Works Cited

Ahmed, M., Kashem, M. A., Rahman, M., & Khatun, S. (2020). Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT). In A. Kasruddin Nasir et al. (Eds.), InECCE2019. Lecture Notes in Electrical Engineering, Vol. 632. Springer, Singapore.

Amore, A. D., Britt, A., Arconada Alvarez, S. J., & Greenleaf, M. N. (2023). A Web-Based Intervention to Address Risk Factors for Maternal Morbidity and Mortality (MAMA LOVE): Development and Evaluation Study. JMIR pediatrics and parenting, 6, e44615. https://doi.org/10.2196/44615

Centers for Disease Control and Prevention. (2019, September 5). *Racial/ethnic disparities in pregnancy-related deaths - United States, 2007–2016*. Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/68/wr/mm6835a3.htm

CSAFRIT2. (n.d.). Maternal Health Risk Data. Retrieved from Kaggle: https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data

World Health Organization. (n.d.). *Maternal Mortality*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality#:~:text=Every%20day%20in%202020%2C%20almost,dropped%20by%20about%2034%25%20worldwide.

[Under publication in IEEE] IoT based Risk Level Prediction Model for Maternal Health Care in the Context of Bangladesh, STI-2020.