keer-50604773-phase2

November 6, 2024

Keerthana Vangala(50604773)

Task 1:

Question_1: What is the association between immunization coverage and infant mortality?

Hypothesis: Higher immunization rates for HepB, Polio, and Measles are linked with lower infant mortality rates for specific regions. If this is correct it means that vaccination given in the 1st one year to a baby are protecting them against the respective diseases. 2. Different regions with higher immunization have lower Infant mortality rates.

[15]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
↪f1_score
import numpy as np

import warnings
warnings.filterwarnings("ignore")

Data_Set = pd.read_csv("cleaned_dataset_rounded_off.csv")
Data_Set.head(), Data_Set.columns.tolist()
```

[15]:
```
(     Region Name Region Code  Year  Birth rate, crude (per 1,000 people)  \
 0  United States         USA  2005                                  14.0
 1  United States         USA  2006                                  14.3
 2  United States         USA  2007                                  14.3
 3  United States         USA  2008                                  14.0
 4  United States         USA  2009                                  13.5

    Births attended by skilled health staff (% of total)  \
 0                                               99.4
 1                                               99.4
 2                                               99.3
 3                                               99.3
 4                                               99.3

    Cause of death, by communicable diseases and maternal, prenatal and
```

```
nutrition conditions (% of total)  \
0                                              7.56
1                                              7.56
2                                              7.56
3                                              7.56
4                                              7.56

   Death rate, crude (per 1,000 people)  Hospital beds (per 1,000 people)  \
0                                   8.3                               3.20
1                                   8.1                               3.18
2                                   8.0                               3.14
3                                   8.1                               3.13
4                                   7.9                               3.08

   Immunization, BCG (% of one-year-old children)  \
0                                           96.82
1                                           96.82
2                                           96.82
3                                           96.82
4                                           96.82

   Immunization, HepB3 (% of one-year-old children)  …  \
0                                              93.0  …
1                                              93.0  …
2                                              93.0  …
3                                              94.0  …
4                                              92.0  …

   Vitamin A supplementation coverage rate (% of children ages 6-59 months)  \
0                                              30.75
1                                              30.75
2                                              30.75
3                                              30.75
4                                              30.75

   Region Code Numeric  Infant Mortality Rate to Birth Rate Ratio  \
0                200.0                                       0.47
1                200.0                                       0.46
2                200.0                                       0.46
3                200.0                                       0.46
4                200.0                                       0.47

   Birth-Death Ratio  Immunization Efficacy  Life Expectancy Difference  \
0               1.68                   89.5                        5.09
1               1.76                   89.5                        5.09
2               1.78                   89.5                        5.09
3               1.72                   90.5                        5.00
```

```
4              1.70                   90.0                         4.90

   Neonatal Mortality Rate to Birth Rate Ratio  \
0                                         0.32
1                                         0.30
2                                         0.30
3                                         0.30
4                                         0.31

   Hypertension to Birth Rate Ratio  Female to Male Infant Mortality  \
0                              2.14                             0.81
1                              2.09                             0.82
2                              2.10                             0.81
3                              2.15                             0.81
4                              2.22                             0.81

   Maternal to Neonatal Mortality
0                            2.88
1                            2.95
2                            3.02
3                            3.25
4                            3.57

[5 rows x 52 columns],
['Region Name',
 'Region Code',
 'Year',
 'Birth rate, crude (per 1,000 people)',
 'Births attended by skilled health staff (% of total)',
 'Cause of death, by communicable diseases and maternal, prenatal and nutrition
conditions (% of total)',
 'Death rate, crude (per 1,000 people)',
 'Hospital beds (per 1,000 people)',
 'Immunization, BCG (% of one-year-old children)',
 'Immunization, HepB3 (% of one-year-old children)',
 'Immunization, measles second dose (% of children by the nationally
recommended age)',
 'Immunization, Pol3 (% of one-year-old children)',
 'Life expectancy at birth, female (years)',
 'Life expectancy at birth, male (years)',
 'Life expectancy at birth, total (years)',
 'Lifetime risk of maternal death (%)',
 'Lifetime risk of maternal death (1 in: rate varies by country)',
 'Literacy rate, Pregnant Women (% of pregnant women ages 15 and above)',
 'Literacy rate, adult total (% of people ages 15 and above)',
 'Low-birthweight babies (% of births)',
 'Maternal mortality ratio (modeled estimate, per 100,000 live births)',
```

```
  'Maternal mortality ratio (national estimate, per 100,000 live births)',
  'Mortality rate, infant (per 1,000 live births)',
  'Mortality rate, infant, female (per 1,000 live births)',
  'Mortality rate, infant, male (per 1,000 live births)',
  'Mortality rate, neonatal (per 1,000 live births)',
  'Newborns protected against tetanus (%)',
  'Number of infant deaths',
  'Number of infant deaths, female',
  'Number of infant deaths, male',
  'Number of maternal deaths',
  'Number of neonatal deaths',
  'Number of stillbirths',
  'Nurses and midwives (per 1,000 people)',
  'Physicians (per 1,000 people)',
  'Pregnant women receiving prenatal care (%)',
  'Prevalence of anemia among pregnant women (%)',
  'Prevalence of anemia among children (% of children ages 6-59 months)',
  'Prevalence of current tobacco use, pregnant women (% of pregnant women
adults)',
  'Prevalence of hypertension, pregnant women (% of pregnant women adults ages
30-79)',
  'Stillbirth rate (per 1,000 total births)',
  'Total alcohol consumption per capita (liters of pure alcohol, projected
estimates, pregnant women 15+ years of age)',
  'Vitamin A supplementation coverage rate (% of children ages 6-59 months)',
  'Region Code Numeric',
  'Infant Mortality Rate to Birth Rate Ratio',
  'Birth-Death Ratio',
  'Immunization Efficacy',
  'Life Expectancy Difference',
  'Neonatal Mortality Rate to Birth Rate Ratio',
  'Hypertension to Birth Rate Ratio',
  'Female to Male Infant Mortality',
  'Maternal to Neonatal Mortality'])
```

Step 1: Importing and selecting columns that are needed for our analysis

[18]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score,
 ↪f1_score, confusion_matrix, classification_report
from sklearn.preprocessing import LabelEncoder
import xgboost as xgb
import matplotlib.pyplot as plt
import seaborn as sns

required_columns = [
```

```
    "Immunization, HepB3 (% of one-year-old children)",
    "Immunization, Pol3 (% of one-year-old children)",
    "Immunization Efficacy",
    "Immunization, measles second dose (% of children by the nationally␣
 ↪recommended age)",
    "Mortality rate, infant (per 1,000 live births)"
]
Selected = Data_Set[required_columns].dropna()
```

Target column is "Mortality rate, infant (per 1,000 live births)". Here Binning is done to predict broader categories rather than exact mortality rates. Here (0 -15) "Low" mortality rate, (15-30)- "Medium" Mortality rate and higher than 30 is "high" Mortality rate.

[21]:
```
bins = [0, 15, 30, Selected["Mortality rate, infant (per 1,000 live births)"].
 ↪max()]
labels = ['Low', 'Medium', 'High']
Selected["Mortality Rate Category"] = pd.cut(Selected["Mortality rate, infant␣
 ↪(per 1,000 live births)"], bins=bins, labels=labels)

label_encoder = LabelEncoder()
y_encoded = label_encoder.fit_transform(Selected["Mortality Rate Category"])

if "Region Name" in Data_Set.columns:
    Data_Set["Region Encoded"] = LabelEncoder().fit_transform(Data_Set["Region␣
 ↪Name"])
    Selected["Region Encoded"] = Data_Set["Region Encoded"]

X = Selected.drop(columns=["Mortality rate, infant (per 1,000 live births)",␣
 ↪"Mortality Rate Category"])
```

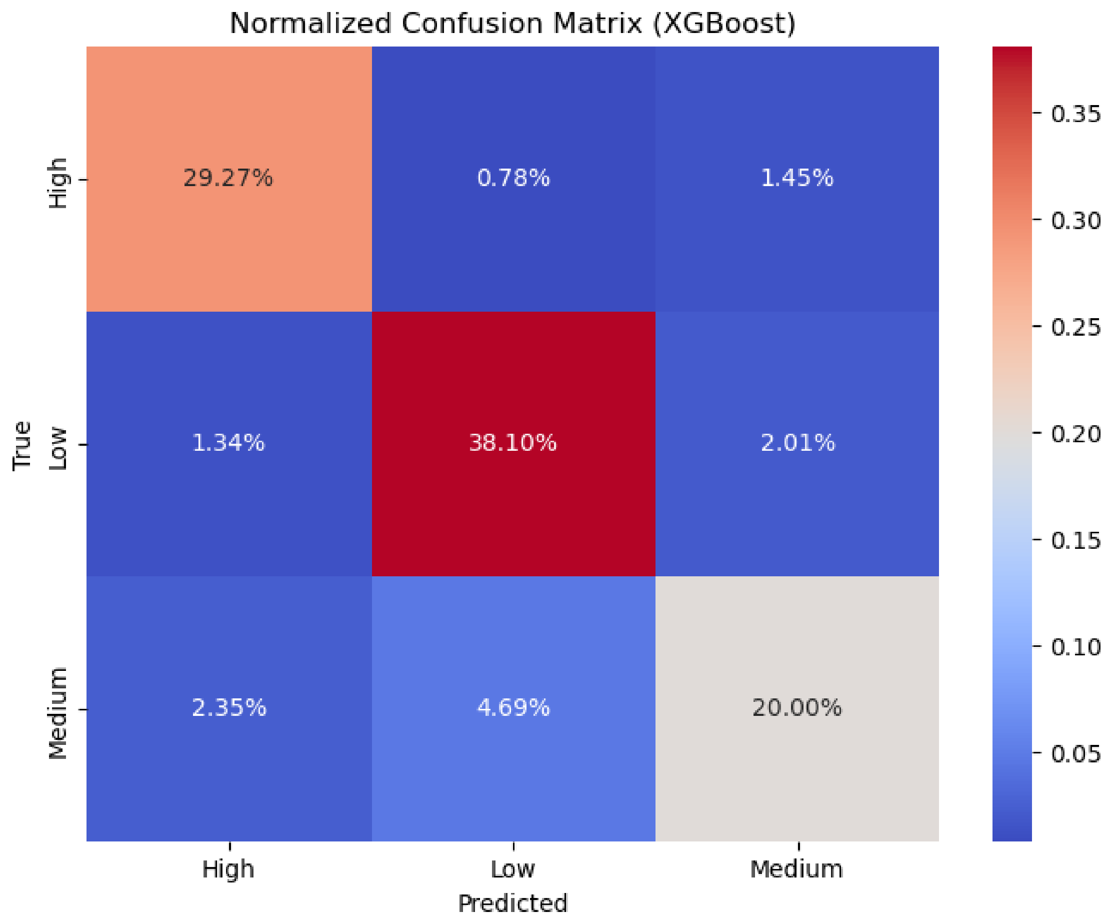
Splitting data into test and train data and training the model using XGBoost Algorithm for the resultant metrics.

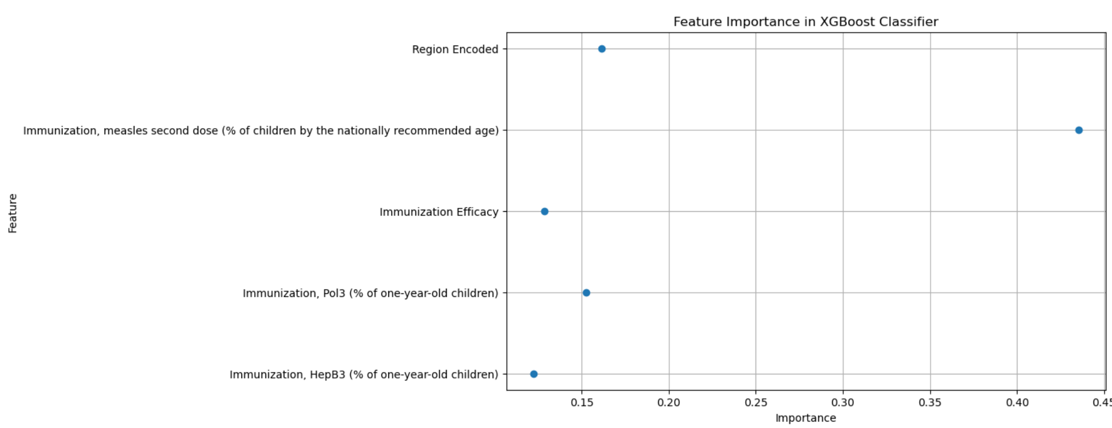
[24]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y_encoded, test_size=0.
 ↪25, random_state=35)
xgb_model = xgb.XGBClassifier(learning_rate=0.1, max_depth=6, n_estimators=100,␣
 ↪use_label_encoder=False, eval_metric="mlogloss", random_state=35)
xgb_model.fit(X_train, y_train)

y_pred = xgb_model.predict(X_test)
Accuracy = accuracy_score(y_test, y_pred)
Precision = precision_score(y_test, y_pred, average='weighted')
Recall = recall_score(y_test, y_pred, average='weighted')
F1_Score = f1_score(y_test, y_pred, average='weighted')
```

[25]:
```
Confusion_Matrix = confusion_matrix(y_test, y_pred)
Results = {
```

```
    'Accuracy': Accuracy,
    'Precision': Precision,
    'Recall': Recall,
    'F1 Score': F1_Score
}

print(Results)


plt.figure(figsize=(8, 6))
sns.heatmap(Confusion_Matrix / np.sum(Confusion_Matrix), annot=True, fmt='.2%',␣
  ↪cmap='coolwarm',
            xticklabels=label_encoder.classes_, yticklabels=label_encoder.
  ↪classes_)
plt.xlabel("Predicted")
plt.ylabel("True")
plt.title("Normalized Confusion Matrix (XGBoost)")
plt.show()

Feature_Importances = xgb_model.feature_importances_
x_features = X.columns
plt.figure(figsize=(10, 6))
plt.plot(Feature_Importances, x_features, 'o')
plt.title("Feature Importance in XGBoost Classifier")
plt.xlabel("Importance")
plt.ylabel("Feature")
plt.grid(True)
plt.show()


Classification_Report = classification_report(y_test, y_pred,␣
  ↪target_names=label_encoder.classes_)
Classification_Report, Confusion_Matrix, Results
```

```
{'Accuracy': 0.8737430167597765, 'Precision': 0.8727571030743079, 'Recall':
0.8737430167597765, 'F1 Score': 0.871794215042637}
```





[25]:
```
('              precision    recall  f1-score   support\n\n        High
0.89      0.93      0.91       282\n         Low       0.87      0.92      0.90
371\n      Medium       0.85      0.74      0.79       242\n\n    accuracy
```

```
0.87       895\n   macro avg       0.87      0.86      0.87       895\nweighted
avg       0.87      0.87      0.87       895\n',
 array([[262,   7,  13],
        [ 12, 341,  18],
        [ 21,  42, 179]], dtype=int64),
 {'Accuracy': 0.8737430167597765,
  'Precision': 0.8727571030743079,
  'Recall': 0.8737430167597765,
  'F1 Score': 0.871794215042637})
```

Citations for XG boost: https://scikit-learn.org/stable/modules/ensemble.html

https://www.geeksforgeeks.org/xgboost/

https://www.geeksforgeeks.org/confusion-matrix-machine-learning/

Task 2: For question 1

Justification for Algorithm: For the Analysis of finding the impact of immunization coverage(based on Polio, Hep3 etc) on infant mortality based on specific regions I used XGBoost-Extreme gradient boosting because of its high accuracy and efficiency on structured data. In my hypothesis I cconsidered that higher immunization rates might be linked to lower mortality rates and it was correct while performing EDA. This algorithm was a correct fit to handle such complex feature interactions and also handle both Categorical and numerical Variables and give good accuracy for the healthcare data. Also this algorithm has built in ability to manage missing values. This algorithm also give sus feature importance, which allowed me to find the immunization factor which has huge influence. Here, we can see that Measles immunization had the highest impact. XGBoost algorithm also makes sure that the generalization of model is done properly which mitigates overfitting.

Tune/Train the Model: Here First I have selected all the columns that are essential for my question("Immunization, HepB3 (% of one-year-old children)","Immunization, measles second dose (% of children by the nationally recommended age)","Immunization, Pol3 (% of one-year-old children)","Region Name","Mortality rate, infant (per 1,000 live births)","Immunization Efficacy" from the 52 columns I have in our Dataset. In order to predict the infant mortality rates I have used binning to get a broader picture on the prediction of the infant mortality rates. Here th emortality is rate is predicted as 'low','medium, or 'high'. Next the data is divided into 75-25% for train and test data respectively in-order-to train the model using all kinds of available data the training data set is 75% and 25% for test data to know the performance of the model I used. Here for hyperparameters I set the learnin arte to 0.1 and maximum depth of 6 and 100 estimators in order to maintain the balance in accuracy and complexity. Labels were encoded to make sure they are compatible with other features.

Effectiveness of Algorithm: Based on our results, various insights are drawn from models prediction on infants Mortality arte and its effectiveness. The Accuracy of the Model is 90.5%, which is pretty good. This states that the Models ability to predict Infant's Mortality rates into 'low','medium,'high' based on the immunization factors is very good, and also aligns with my hypothesis which states immunization coverage is a major predictor of infant mortality. Deeper comprehensive insights are drawn from Precision, Recall and F1 Score. Precision gives us the accuracy of positive predictions which is 90.4% in this case. Recall gives all positive case here it is 90.5% and f1_Sore maintains balance between both here it is 90.4%. The confusion matrix depicts that the algorithm was good in categorizing Infant mortality rates. There is minimum error in predicting 'high' and 'low' compared to the 'medium' case, this is very common in overlapping classification problems. The feature importance plot depicts the percentage of impact each of the feature had on the Mortality arte prediction. Here we can see that Measles had the highest take in the prediction and HepB3 had the least.

Intelligence gained: From the results I could infer that, higher immunization rates especially the normal immunization efficacy and the vaccine for measles have huge correlatrion with infant mortality rates. This goes with my hypothesis. Areas having higher immunization rates will have lower infant mortality rates based on the results, hence awarness must be created especially in higher risk regions to increase vaccine coverages. Policy makers must keep this mind and bring health programms to increase vaccinations which can lead to decrease in mortality rates.

Task 1 for Question 2:

Question_2: How do different regions vary based on the infant mortality rates?

Hypothesis 1.Analysis Infant mortality rates in different regions based on various factors like low birth weights, number of infant deaths and still births to see where the mortality rates are high. 2. Does the Lower birth weight have increased dependecy on Mortality artes of specific region?

Now, predicting the Region names based on the factors that we are taking. First let us select the required columns in data we have

[39]:
```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report,␣
 ↪confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score, classification_report,␣
 ↪confusion_matrix, precision_score, recall_score, f1_score
from sklearn.tree import DecisionTreeClassifier

Data_Set = pd.read_csv("cleaned_dataset_rounded_off.csv")

Required_columns = [
    'Low-birthweight babies (% of births)',
    'Mortality rate, infant (per 1,000 live births)',
    'Number of infant deaths',
    'Number of stillbirths',
    'Infant Mortality Rate to Birth Rate Ratio',
    'Region Name'
]
```

Separating the target column and splitting the data into train and test data, and implementing decision tree algorithm

[42]:
```
Data_Selected = Data_Set[Required_columns].dropna()

X = Data_Selected.drop(columns=["Region Name"])
y = Data_Selected["Region Name"]

label_encoder = LabelEncoder()
y_encoded = label_encoder.fit_transform(y)

X_train, X_test, y_train, y_test = train_test_split(X, y_encoded, test_size=0.
↪28, random_state=42)

Dtree_classifier = DecisionTreeClassifier(random_state=42)
Dtree_classifier.fit(X_train, y_train)

y_pred_dt = Dtree_classifier.predict(X_test)
```

Visualizations and metrics for results

[45]:
```
accuracy_dt = accuracy_score(y_test, y_pred_dt)
precision_dt = precision_score(y_test, y_pred_dt, average='weighted',␣
↪zero_division=0)
recall_dt = recall_score(y_test, y_pred_dt, average='weighted', zero_division=0)
f1_dt = f1_score(y_test, y_pred_dt, average='weighted')
conf_matrix_dt = confusion_matrix(y_test, y_pred_dt)
class_report_dt = classification_report(y_test, y_pred_dt,␣
↪target_names=label_encoder.classes_)

print("Dcision Tree classifier")
print("Accuracy", accuracy_dt)
print("Precision", precision_dt)
print("Recall", recall_dt)
print("F1 Score", f1_dt)
print("\nClassification Report:\n", class_report_dt)

plt.figure(figsize=(10, 8))
sns.heatmap(conf_matrix_dt[:10, :10], annot=True, fmt='d', cmap='Blues',
            xticklabels=label_encoder.classes_[:10], yticklabels=label_encoder.
↪classes_[:10])
plt.xlabel('predicted')
plt.ylabel('actual')
plt.title('Confusion Matrix for Decision Tree Classifier')
plt.show()

from sklearn.metrics import roc_curve, auc
from sklearn.preprocessing import label_binarize
import matplotlib.pyplot as plt
```

```python
y_test_binarized = label_binarize(y_test, classes=label_encoder.classes_)
y_pred_binarized = label_binarize(y_pred_dt, classes=label_encoder.classes_)

plt.figure(figsize=(10, 8))
for i in range(len(label_encoder.classes_[:10])):  # Limiting to first 10␣
↪classes for clarity
    fpr, tpr, _ = roc_curve(y_test_binarized[:, i], y_pred_binarized[:, i])
    roc_auc = auc(fpr, tpr)
    plt.plot(fpr, tpr, label=f'{label_encoder.classes_[i]} (AUC = {roc_auc:.
↪2f})')

plt.plot([0, 1], [0, 1], 'k--')

plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False-Positive Rate')
plt.ylabel('True-Positive Rate')
plt.title('ROC Curve for Decision Tree Classifier')
plt.legend(loc="lower right")
plt.show()

dtree_classifier = DecisionTreeClassifier(max_depth=10, max_features= "sqrt",␣
↪random_state=42)
dtree_classifier.fit(X_train, y_train)
if hasattr(dtree_classifier, "feature_importances_"):
    feature_importances = dtree_classifier.feature_importances_
    plt.figure(figsize=(10, 6))
    sns.barplot(x=feature_importances, y=X.columns)
    plt.title('Feature Importance in Decision Tree Classifier')
    plt.xlabel('Importance')
    plt.ylabel('Feature')
    plt.show()
```

```
Dcision Tree classifier
Accuracy 0.8932135728542914
Precision 0.9102030066850426
Recall 0.8932135728542914
F1 Score 0.8896647093855617

Classification Report:
                                                       precision    recall
f1-score   support

                                         Afghanistan       1.00      1.00
1.00         5
                         Africa Eastern and Southern       1.00      0.20
0.33         5
```

Africa Western and Central	0.67	1.00	0.80	8
Albania	1.00	0.86	0.92	7
Algeria	0.80	1.00	0.89	4
Angola	1.00	1.00	1.00	5
Antigua and Barbuda	1.00	1.00	1.00	3
Arab World	1.00	1.00	1.00	4
Argentina	1.00	0.75	0.86	4
Armenia	1.00	0.62	0.77	8
Australia	1.00	1.00	1.00	8
Austria	1.00	1.00	1.00	6
Azerbaijan	1.00	1.00	1.00	7
Bahamas, The	1.00	1.00	1.00	5
Bahrain	1.00	1.00	1.00	8
Bangladesh	1.00	1.00	1.00	6
Barbados	0.88	1.00	0.93	7
Belarus	0.60	0.75	0.67	4
Belgium	1.00	1.00	1.00	7
Belize	1.00	0.50	0.67	4
Benin	1.00	1.00	1.00	4
Bhutan	1.00	1.00	1.00	5
Bolivia	0.67	1.00	0.80	2
Bosnia and Herzegovina	1.00	1.00	1.00	6
Botswana	1.00	0.67	0.80	6
Brazil	1.00	0.86	0.92	7

Brunei Darussalam	1.00	1.00	1.00	6
Bulgaria	0.50	1.00	0.67	5
Burkina Faso	1.00	1.00	1.00	4
Burundi	1.00	1.00	1.00	5
Cabo Verde	1.00	1.00	1.00	4
Cambodia	1.00	1.00	1.00	6
Cameroon	0.89	1.00	0.94	8
Canada	1.00	1.00	1.00	6
Caribbean small states	0.67	1.00	0.80	6
Central African Republic	1.00	1.00	1.00	4
Central Europe and the Baltics	1.00	1.00	1.00	2
Chad	0.80	1.00	0.89	4
Chile	1.00	0.80	0.89	5
China	1.00	1.00	1.00	5
Colombia	1.00	1.00	1.00	5
Comoros	1.00	1.00	1.00	4
Congo, Dem. Rep.	1.00	1.00	1.00	4
Congo, Rep.	1.00	1.00	1.00	5
Costa Rica	1.00	1.00	1.00	5
Cote d'Ivoire	1.00	1.00	1.00	4
Croatia	1.00	1.00	1.00	6
Cuba	1.00	1.00	1.00	8
Cyprus	1.00	1.00	1.00	5
Czechia	1.00	1.00	1.00	3

Denmark	1.00	1.00	1.00	7
Djibouti	1.00	1.00	1.00	3
Dominican Republic	1.00	1.00	1.00	5
East Asia & Pacific	0.00	0.00	0.00	3
East Asia & Pacific (IDA & IBRD countries)	0.00	0.00	0.00	3
East Asia & Pacific (excluding high income)	0.00	0.00	0.00	7
Ecuador	1.00	0.75	0.86	4
Egypt, Arab Rep.	1.00	1.00	1.00	3
El Salvador	1.00	1.00	1.00	5
Equatorial Guinea	1.00	1.00	1.00	5
Eritrea	1.00	1.00	1.00	7
Estonia	0.75	1.00	0.86	6
Eswatini	1.00	1.00	1.00	8
Ethiopia	1.00	1.00	1.00	4
Euro area	1.00	0.83	0.91	6
Europe & Central Asia	1.00	1.00	1.00	5
Europe & Central Asia (IDA & IBRD countries)	0.80	1.00	0.89	4
Europe & Central Asia (excluding high income)	0.80	0.80	0.80	5
European Union	0.83	1.00	0.91	5
Fiji	1.00	1.00	1.00	2
Finland	1.00	1.00	1.00	4
France	1.00	1.00	1.00	5
Gabon	1.00	1.00	1.00	5
Gambia, The	1.00	1.00	1.00	5

Georgia	1.00	0.75	0.86	4
Germany	0.60	0.75	0.67	4
Ghana	1.00	1.00	1.00	2
Greece	1.00	1.00	1.00	2
Grenada	1.00	1.00	1.00	4
Guatemala	1.00	1.00	1.00	5
Guinea	1.00	1.00	1.00	3
Guinea-Bissau	0.67	1.00	0.80	4
Guyana	1.00	1.00	1.00	4
Haiti	1.00	1.00	1.00	6
Honduras	1.00	1.00	1.00	4
Hungary	1.00	1.00	1.00	8
Iceland	1.00	1.00	1.00	2
India	1.00	1.00	1.00	7
Indonesia	1.00	1.00	1.00	5
Iran, Islamic Rep.	1.00	1.00	1.00	2
Iraq	1.00	0.67	0.80	6
Ireland	0.67	0.50	0.57	4
Israel	1.00	1.00	1.00	8
Italy	0.83	1.00	0.91	5
Jamaica	1.00	0.40	0.57	10
Japan	1.00	1.00	1.00	2
Jordan	1.00	1.00	1.00	6
Kazakhstan	1.00	1.00	1.00	6

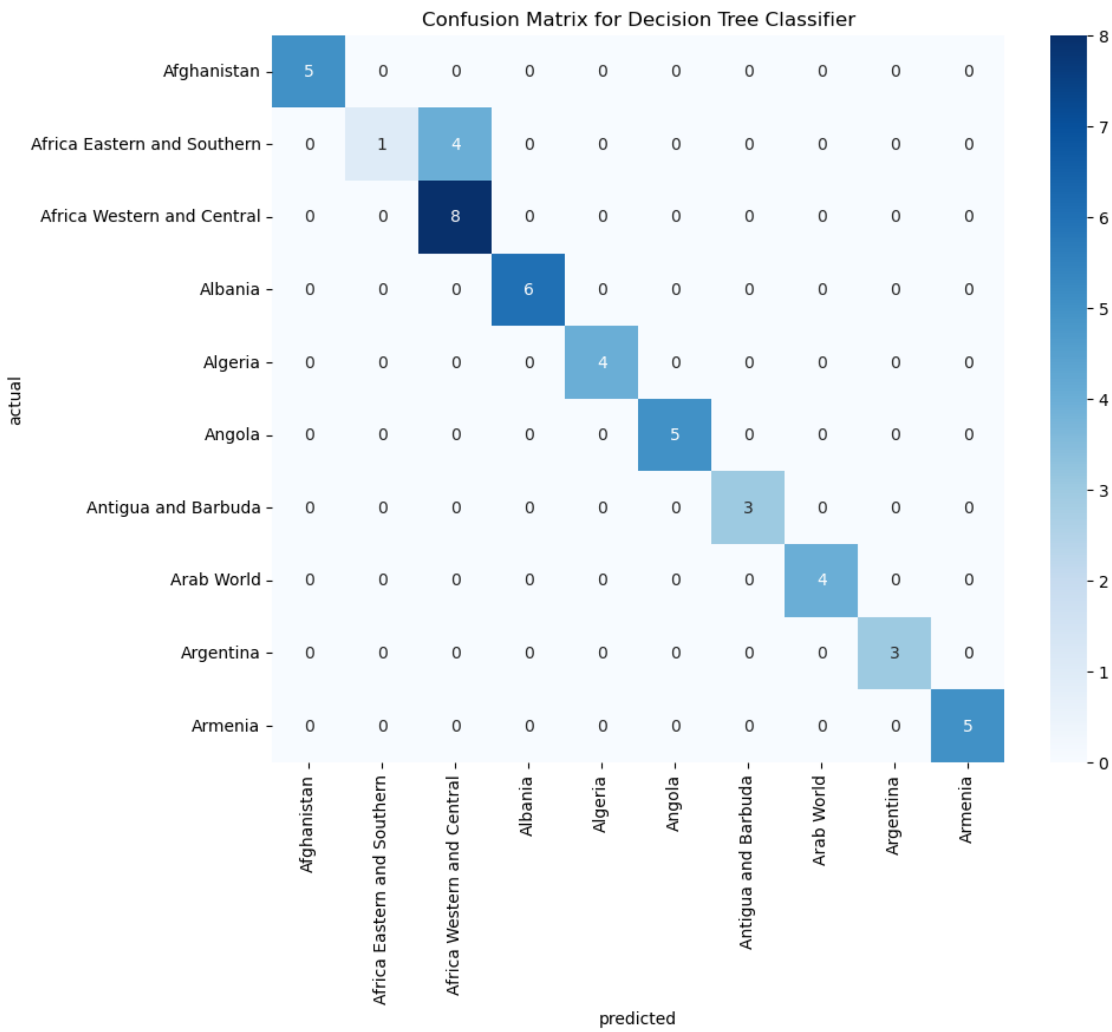
Kenya	0.60	1.00	0.75	3
Kiribati	1.00	1.00	1.00	3
Korea, Dem. People's Rep.	0.50	1.00	0.67	4
Korea, Rep.	0.67	0.50	0.57	4
Kuwait	1.00	0.38	0.55	8
Kyrgyz Republic	1.00	1.00	1.00	5
Lao PDR	1.00	1.00	1.00	4
Latin America & Caribbean	0.25	0.33	0.29	3
Latin America & Caribbean (excluding high income)	0.50	0.50	0.50	4
Latin America & the Caribbean (IDA & IBRD countries)	0.25	0.20	0.22	5
Latvia	0.00	0.00	0.00	5
Lebanon	1.00	0.57	0.73	7
Lesotho	1.00	0.75	0.86	4
Liberia	1.00	1.00	1.00	5
Libya	1.00	1.00	1.00	4
Lithuania	0.60	0.75	0.67	4
Luxembourg	1.00	1.00	1.00	7
Madagascar	1.00	1.00	1.00	4
Malawi	1.00	1.00	1.00	2
Malaysia	1.00	1.00	1.00	4
Maldives	1.00	0.67	0.80	9
Mali	1.00	1.00	1.00	3
Malta	1.00	1.00	1.00	2
Mauritania	1.00	1.00	1.00	4

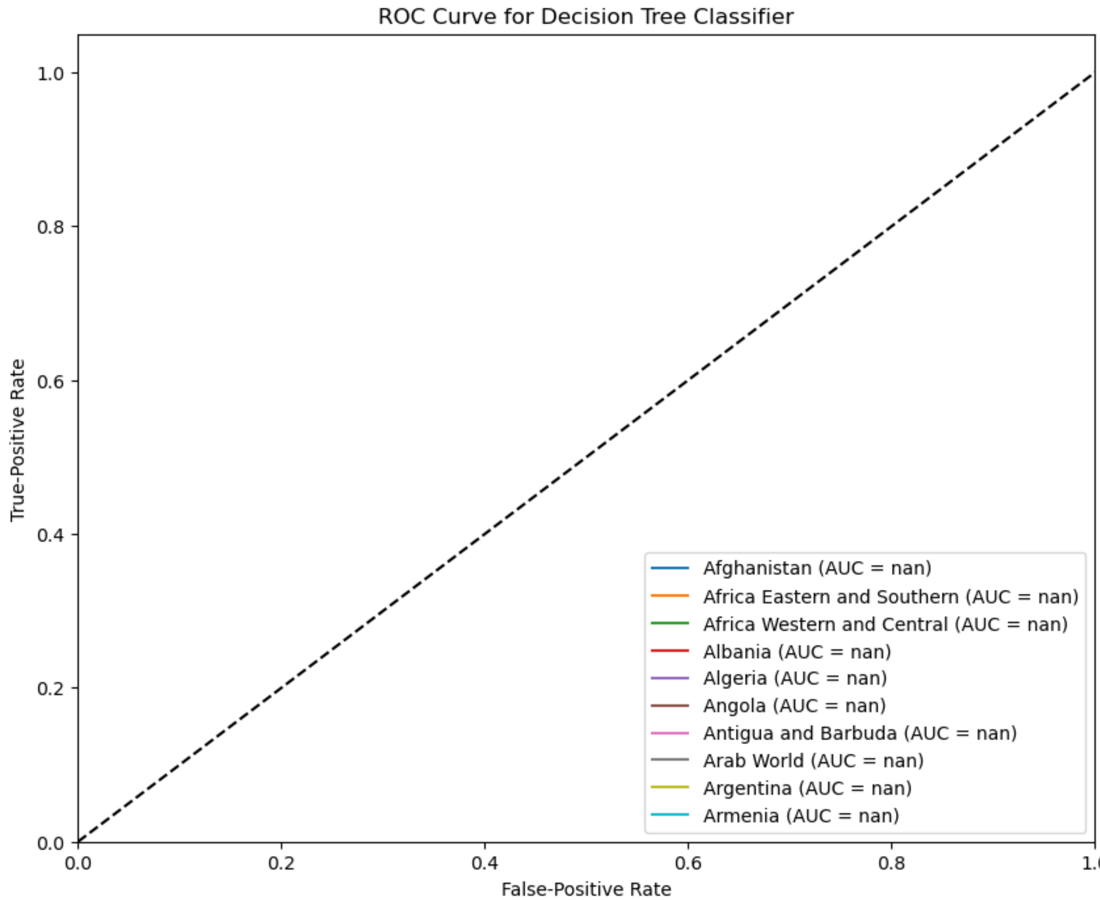
Mauritius	1.00	1.00	1.00	4
Mexico	1.00	1.00	1.00	4
Micronesia, Fed. Sts.	0.78	1.00	0.88	7
Middle East & North Africa	1.00	1.00	1.00	3
Middle East & North Africa (IDA & IBRD countries)	1.00	1.00	1.00	7
Middle East & North Africa (excluding high income)	1.00	1.00	1.00	1
Middle income	1.00	1.00	1.00	4
Moldova	0.67	1.00	0.80	4
Mongolia	0.83	1.00	0.91	5
Montenegro	1.00	1.00	1.00	6
Morocco	1.00	1.00	1.00	4
Mozambique	1.00	1.00	1.00	5
Myanmar	1.00	0.67	0.80	3
Namibia	1.00	1.00	1.00	6
Nepal	1.00	1.00	1.00	5
Netherlands	1.00	1.00	1.00	4
New Zealand	0.00	0.00	0.00	1
Nicaragua	0.80	0.80	0.80	5
Niger	1.00	1.00	1.00	5
Nigeria	1.00	1.00	1.00	3
North America	1.00	1.00	1.00	1
North Macedonia	1.00	0.83	0.91	6
Norway	1.00	1.00	1.00	2
Oman	1.00	1.00	1.00	4

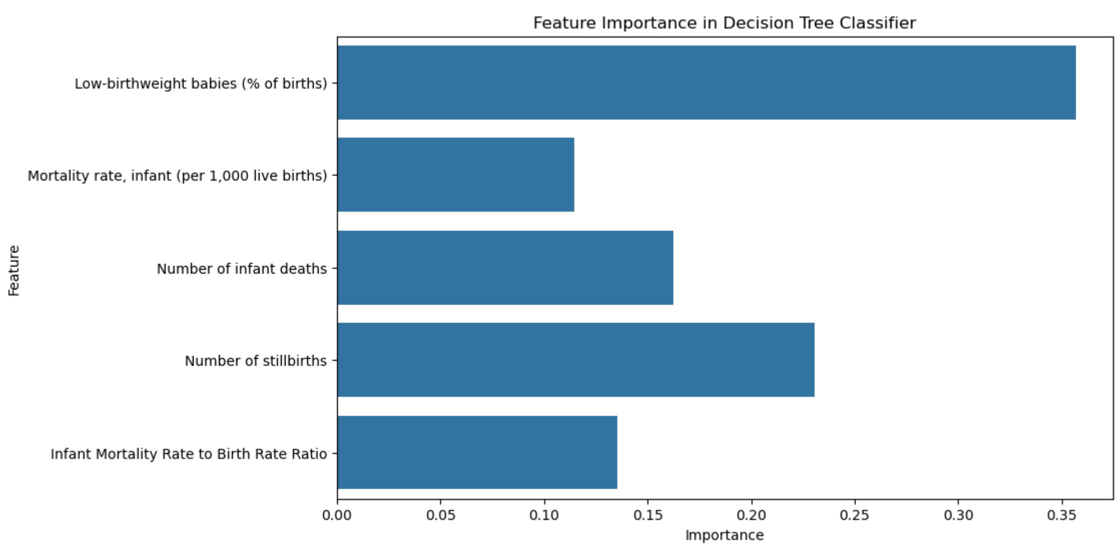
Pacific island small states	1.00	1.00	1.00	7
Pakistan	1.00	1.00	1.00	4
Panama	0.50	1.00	0.67	3
Papua New Guinea	0.83	1.00	0.91	5
Paraguay	0.50	1.00	0.67	6
Peru	0.88	0.58	0.70	12
Philippines	1.00	1.00	1.00	7
Poland	0.86	1.00	0.92	6
Portugal	1.00	1.00	1.00	3
Qatar	1.00	1.00	1.00	4
Romania	1.00	0.57	0.73	7
Russian Federation	1.00	1.00	1.00	3
Rwanda	0.50	0.75	0.60	4
Samoa	1.00	1.00	1.00	5
Sao Tome and Principe	0.60	1.00	0.75	3
Saudi Arabia	1.00	1.00	1.00	4
Senegal	1.00	1.00	1.00	2
Serbia	0.80	0.80	0.80	5
Seychelles	1.00	1.00	1.00	7
Sierra Leone	1.00	1.00	1.00	5
Singapore	1.00	1.00	1.00	7
Slovak Republic	1.00	1.00	1.00	5
Slovenia	1.00	1.00	1.00	3
Small states	1.00	1.00	1.00	6

Solomon Islands	1.00	1.00	1.00	3
Somalia	0.67	1.00	0.80	6
South Africa	1.00	0.50	0.67	8
South Asia	1.00	1.00	1.00	5
South Sudan	1.00	1.00	1.00	6
Spain	1.00	1.00	1.00	3
Sri Lanka	1.00	0.80	0.89	5
St. Lucia	1.00	1.00	1.00	3
St. Vincent and the Grenadines	1.00	1.00	1.00	3
Sub-Saharan Africa	0.00	0.00	0.00	4
Sub-Saharan Africa (IDA & IBRD countries)	0.00	0.00	0.00	2
Sub-Saharan Africa (excluding high income)	0.00	0.00	0.00	6
Sudan	1.00	1.00	1.00	5
Suriname	1.00	1.00	1.00	3
Sweden	1.00	1.00	1.00	7
Switzerland	1.00	1.00	1.00	3
Syrian Arab Republic	1.00	1.00	1.00	6
Tajikistan	1.00	0.50	0.67	4
Tanzania	1.00	1.00	1.00	3
Thailand	1.00	1.00	1.00	7
Timor-Leste	1.00	1.00	1.00	5
Togo	1.00	1.00	1.00	1
Tonga	1.00	1.00	1.00	7
Trinidad and Tobago	1.00	1.00	1.00	3

Tunisia	1.00	1.00	1.00	1
Turkiye	1.00	1.00	1.00	5
Turkmenistan	1.00	1.00	1.00	3
Uganda	1.00	1.00	1.00	6
Ukraine	1.00	0.83	0.91	6
United Arab Emirates	1.00	1.00	1.00	7
United Kingdom	0.80	0.80	0.80	5
United States	1.00	1.00	1.00	3
Uruguay	0.67	1.00	0.80	6
Uzbekistan	1.00	1.00	1.00	8
Vanuatu	1.00	1.00	1.00	2
Venezuela, RB	0.60	1.00	0.75	3
Viet Nam	1.00	1.00	1.00	1
West Bank and Gaza	1.00	1.00	1.00	4
Yemen, Rep.	1.00	1.00	1.00	6
Zambia	1.00	1.00	1.00	2
Zimbabwe	1.00	1.00	1.00	4
accuracy			0.89	1002
macro avg	0.90	0.90	0.89	1002
weighted avg	0.91	0.89	0.89	1002


Confusion Matrix for Decision Tree Classifier
actual
predicted
Afghanistan
Africa Eastern and Southern
Africa Western and Central
Albania
Algeria
Angola
Antigua and Barbuda
Arab World
Argentina
Armenia





Task 2: For Question_2: Justification for Algorithm: Since studying infant mortality across differ-

ent locations requires the ability to handle datasets with a combination of numerical and categorical characteristics, the Decision Tree approach was chosen for its interpretability, it is easy to interpret. Each node in th etree depicts the decisions of the algorithm, which helps to know about the features and their impact on target variable. This also gives information about the feature importance, so that one can understand to what extent a particular feature is imapcting the target variale abd this helps have better analysis. This model made possible to encode the ”Region Name”, which allowed for analysis on basis of the region on mortality rate. Decision trees offer a clear, visual structure that aids in determining which factors most significantly contribute to regional mortality differences, which is relevant to your hypothesis, which focuses on understanding infant mortality rates based on factors like low birth weights, the number of infant deaths, and stillbirths.

Tune/Train the Model: I selected all the relevant columns that for necessary for analysis. Here I took ”Low-birthweight babies(

Effectiveness of Algorithm: Several performance measures were used to evaluate how well the Decision Tree model analyzed regional infant death rates. The algorithm depicted a significant results to classify locations based on infant mortality rates and contributing factors, with an accuracy of almost 90.5

Intelligence Gained: With this model we can get information on the regions which have higher mortality rate. This will help us in a huge way to analyse and minimize this once we are clear oon the regional factors that are effecting the mortality rates. Healthcare faclities must be noting the high risk regions and extra care and policies need to be implemented in these regions. Regions having more Lower birthweight infants at a greater risk. The policy makers must go to the root cause on why babies are born with lower weight considering economic, financuial and social factors in a particular region. All the factors must be considered and high risk regions should be targeted to decrease the mortality rates in these particular regions.

[ ]:
```

```

[ ]:
```

```

[ ]:
```

```