Model Development Phase - Initial Model Training Code, Model Validation and Evaluation Report

Date: 30 July 2025

Project Title: AnemiaSense — Machine Learning Based Anemia Detection

Maximum Marks: 10 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training phase for AnemiaSense involved implementing and testing multiple machine learning algorithms to evaluate their effectiveness in detecting anemia. This process included splitting the dataset into training and testing subsets, training models on the training data, and validating performance using accuracy scores and classification reports. By comparing the accuracy of different models, the most suitable algorithm for deployment was identified.

Python Code Implementation

```python
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Load dataset
df = pd.read_csv('anemia.csv')

# Split features and target
X = df.drop('Result', axis=1)
Y = df['Result']

# Train-test split
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=25)
```

```
# Train and evaluate models
models = {
    "Logistic Regression": LogisticRegression(max_iter=1000),
    "Random Forest Classifier": RandomForestClassifier(),
    "Decision Tree Classifier": DecisionTreeClassifier(),
    "Gaussian Naive Bayes": GaussianNB(),
    "Support Vector Machine": SVC(),
    "Gradient Boosting Classifier": GradientBoostingClassifier()
}

for name, model in models.items():
    model.fit(x_train, y_train)
    y_pred = model.predict(x_test)
    acc = accuracy_score(y_test, y_pred)
    print(f"{name}: Accuracy = {acc*100:.2f}%")
```

Model Performance Comparison

The table below summarises the test accuracy scores for each trained model:

Model	Accuracy Score (%)
Logistic Regression	99.65
Random Forest Classifier	100.00
Decision Tree Classifier	100.00
Gaussian Naive Bayes	92.98
Support Vector Machine	90.88
Gradient Boosting Classifier	100.00