Delayed Defect Detection - Source Code

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
import joblib

# Load data
data = pd.read_csv('../data/defect_data.csv')

# Feature Engineering
data['Production_Date'] = pd.to_datetime(data['Production_Date'])
data['Detection_Date'] = pd.to_datetime(data['Detection_Date'])
data['Detection_Delay_Days'] = (data['Detection_Date'] - data['Production_Date']).dt.days

# Label: Delayed Detection (Delay > 30 days)
data['Delayed_Detection'] = (data['Detection_Delay_Days'] > 30).astype(int)

# Encoding categorical features
data = pd.get_dummies(data, columns=['Defect_Type', 'Severity', 'Root_Cause'], drop_first=True)

# Train-Test Split
X = data.drop(columns=['Defect_ID', 'Product_ID', 'Production_Date', 'Detection_Date', 'Detection_Delay_Days',
'Delayed_Detection'])
y = data['Delayed_Detection']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Model Training
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Evaluation
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))

# Save Model
joblib.dump(model, '../models/defect_delay_model.pkl')
```