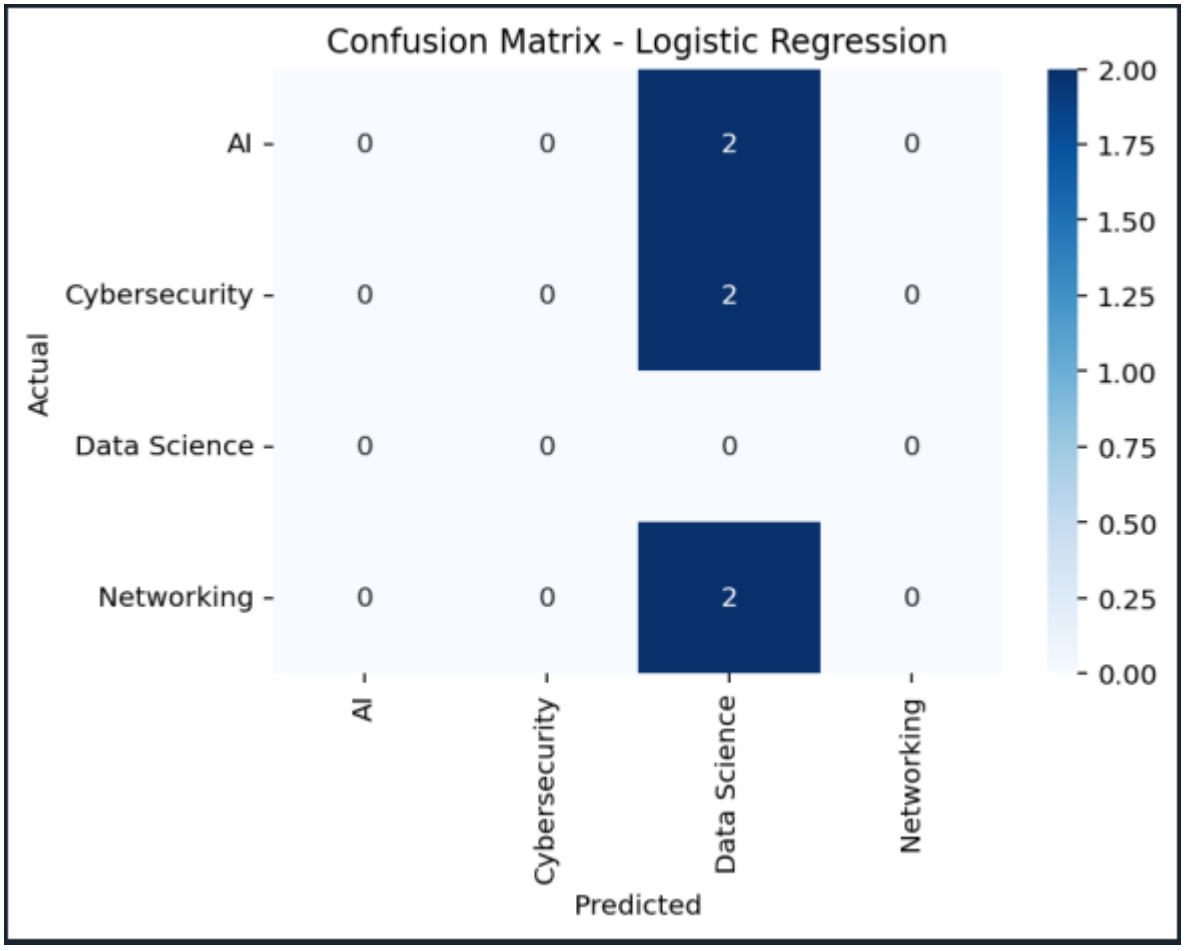
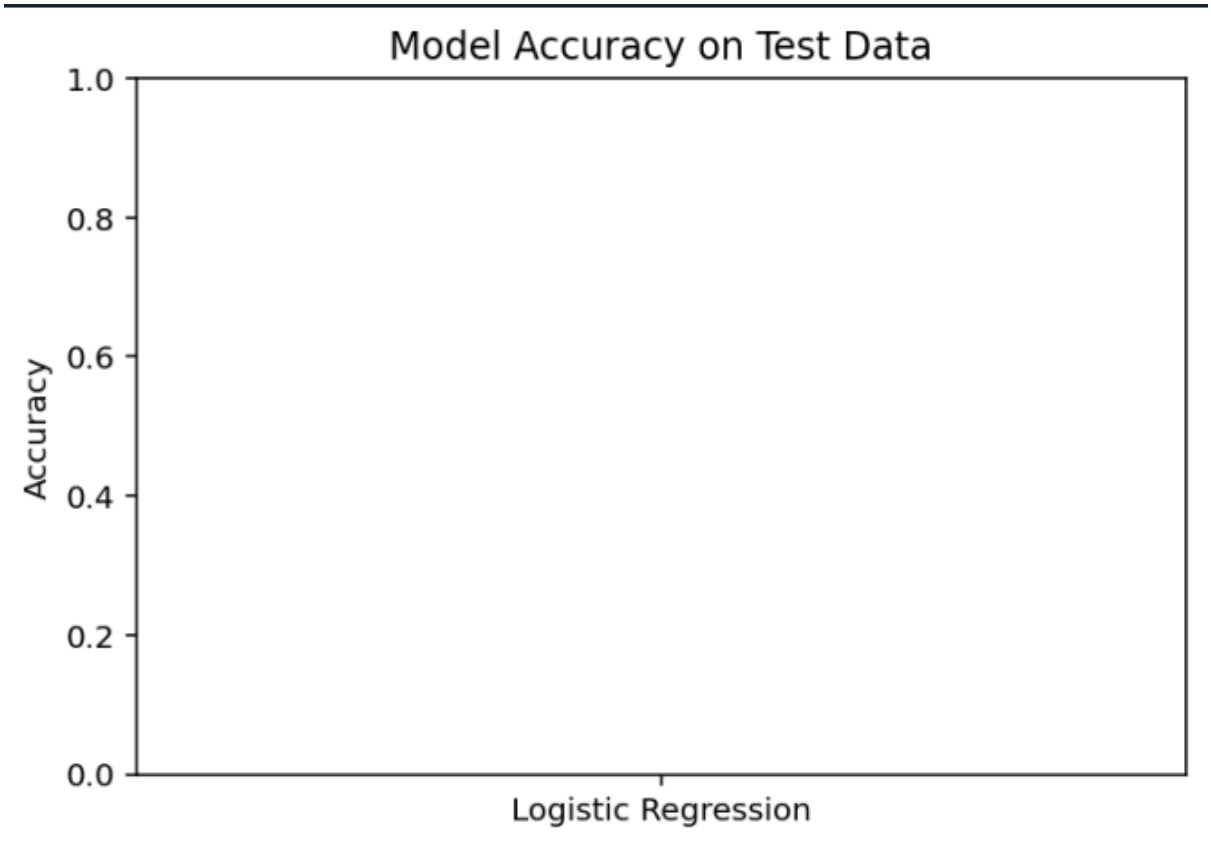
ML Mini Project

https://github.com/tejasholkar/ML-research-paper-model-.git

This is the model which classifies research paper on the basis of their abstract they belong to like This paper presents a new deep learning model... so this paper belongs to AI field. The goal is to read a CSV file of research papers, extract their abstracts, train a machine learning model.

```python
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import seaborn as sns

data = pd.read_csv(r"C:\Users\Tejas Holkar\Desktop\Downloads\research_papers.csv")

print("Sample Data:\n", data.head())

X = data['Abstract']
y = data['Category']

vectorizer = TfidfVectorizer(stop_words='english', max_features=500)
X_tfidf = vectorizer.fit_transform(X)

X_train, X_test, y_train, y_test = train_test_split(
    X_tfidf, y, test_size=0.3, random_state=42
)

model = LogisticRegression(max_iter=200)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print("\nClassification Report:\n", classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))

cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 4))
sns.heatmap(cm, annot=True, cmap='Blues', xticklabels=model.classes_, yticklabels=model.classes_)
plt.title("Confusion Matrix - Logistic Regression")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

accuracy = accuracy_score(y_test, y_pred)
plt.bar(["Logistic Regression"], [accuracy], color='teal')
plt.ylabel("Accuracy")
plt.title("Model Accuracy on Test Data")
plt.ylim(0, 1)
plt.show()
```


Confusion Matrix - Logistic Regression
AI
Cybersecurity
Data Science
Networking
Actual
Predicted
0
0
2
0
0
0
2
0
0
0
0
0
0
0
2
0
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00


Model Accuracy on Test Data
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
Logistic Regression