```
#Lê Hoài Linh
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
from sklearn.neural_network import MLPClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Reshape
from tensorflow.keras.optimizers import Adam
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.layers import Dropout
from tensorflow.keras.preprocessing.sequence import TimeseriesGenerator
from tensorflow.keras.callbacks import History


from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
df = pd.read_csv('/content/drive/MyDrive/5percent_MSSQL.csv')
df = df.applymap(lambda x: x if not isinstance(x, str) else None)
df.replace([np.inf, -np.inf], np.nan, inplace=True)
df = df[~df.isin([np.nan]).any(axis=1)]
df = df.dropna(axis=1, how='all')
```

```
<ipython-input-4-70022ddfaade>:1: DtypeWarning: Columns (86) have mixed types. Specify dtype option on import or set low_memory=Fals
  df = pd.read_csv('/content/drive/MyDrive/5percent_MSSQL.csv')
```

```
print(df)
```

```
        288785       4631911      121994        16055             33609        17
        288786       2610375        9448        32855              6197        17
        288787       1998542       24043        15143             14340        17
        288788       4572056       12645        32253             43549        17

                Flow Duration  Total Fwd Packets  Total Backward Packets  \
        0                   1                  2                       0
        1                  50                  2                       0
        2                   2                  2                       0
        3                  49                  2                       0
        4                   1                  2                       0
        ...               ...                ...                     ...
        288784              1                  2                       0
        288785              1                  2                       0
        288786              1                  2                       0
        288787              1                  2                       0
        288788              1                  2                       0

                Total Length of Fwd Packets  Total Length of Bwd Packets  ...  \
        0                             888.0                          0.0  ...
        1                             862.0                          0.0  ...
        2                            1108.0                          0.0  ...
        3                             952.0                          0.0  ...
        4                            2200.0                          0.0  ...
        ...                             ...                          ...  ...
        288784                        980.0                          0.0  ...
        288785                        888.0                          0.0  ...
        288786                        988.0                          0.0  ...
        288787                       2944.0                          0.0  ...
        288788                        856.0                          0.0  ...

                Active Mean  Active Std  Active Max  Active Min  Idle Mean  Idle Std  \
        0               0.0         0.0         0.0         0.0        0.0       0.0
        1               0.0         0.0         0.0         0.0        0.0       0.0
        2               0.0         0.0         0.0         0.0        0.0       0.0
        3               0.0         0.0         0.0         0.0        0.0       0.0
        4               0.0         0.0         0.0         0.0        0.0       0.0
        ...             ...         ...         ...         ...        ...       ...
```

```
2           0.0       0.0           0.0        1
3           0.0       0.0           0.0        1
4           0.0       0.0           0.0        1
...         ...       ...           ...      ...
288784      0.0       0.0           0.0        1
288785      0.0       0.0           0.0        1
288786      0.0       0.0           0.0        1
288787      0.0       0.0           0.0        1
288788      0.0       0.0           0.0        1

[262869 rows x 84 columns]
```

```
df.head(100)
```

	Unnamed: 0.1	Unnamed: 0	Source Port	Destination Port	Protocol	Flow Duration	Total Fwd Packets	Total Backward Packets
0	568269	42641	55989	50529	17	1	2	0
1	3914458	89976	61850	36734	17	50	2	0
2	2789854	85749	39757	45349	17	2	2	0
3	2834358	86263	11300	20633	17	49	2	0
4	5649515	25038	61850	3134	17	1	2	0
...	...	...	...	...	...	...	...	...
102	3048392	41314	5955	28910	17	2	2	0
103	1037127	20190	55788	22408	17	2	2	0
104	2612643	93692	34863	57068	17	1	2	0
105	2746016	70311	5606	63388	17	1	2	0
106	394004	101238	65378	21023	17	1	2	0

100 rows × 84 columns

```
#Chia dữ liệu thành hai phần
X = df.iloc[:, :-1]
y = df['Flow Duration']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
#chuẩn hóa dữ liệu
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
```

```
#huấn luyện mô hình phân loại
knn_model = KNeighborsClassifier(n_neighbors=5)
knn_model.fit(X_train_scaled, y_train)
```

```
▾ KNeighborsClassifier
KNeighborsClassifier()
```

```
# Chuyển đổi X_test trước khi dự đoán
X_test_scaled = scaler.transform(X_test)
```

```
# Dự đoán với mô hình KNN
y_pred_knn = knn_model.predict(X_test_scaled)
```

```
# Tính toán các chỉ số đánh giá
print("KNN Accuracy:", accuracy_score(y_test, y_pred_knn))
print("KNN Precision:", precision_score(y_test, y_pred_knn, average='weighted'))
print("KNN Recall:", recall_score(y_test, y_pred_knn, average='weighted'))
print("KNN F1 Score:", f1_score(y_test, y_pred_knn, average='weighted'))
```

```
KNN Accuracy: 0.9008445239091566
KNN Precision: 0.8960723430194333
KNN Recall: 0.9008445239091566
KNN F1 Score: 0.8975692405855584
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-defined ar
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Recall is ill-defined and b
  _warn_prf(average, modifier, msg_start, len(result))
```

```
#mô hình Multilayer Perceptron (MLP) từ thư viện scikit-learn để tạo và huấn luyện một mạng nơ-ron
ann_model = MLPClassifier(hidden_layer_sizes=(100,), activation='tanh', max_iter=10)
ann_model.fit(X_train_scaled, y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/neural_network/_multilayer_perceptron
  warnings.warn(
```

```
MLPClassifier
MLPClassifier(activation='tanh', max_iter=10)
```

```
#đánh giá hiệu suất của mô hình Neural Network (DNN) trên tập dữ liệu kiểm thử đã được chuẩn hóa.
y_pred_ann = ann_model.predict(X_test_scaled)
print("DNN Accuracy:", accuracy_score(y_test, y_pred_ann))
print("DNN Precision:", precision_score(y_test, y_pred_ann, average='weighted'))
print("DNN Recall:", recall_score(y_test, y_pred_ann, average='weighted'))
print("DNN F1 Score:", f1_score(y_test, y_pred_ann, average='weighted'))
```

```
DNN Accuracy: 0.9422338037813368
DNN Precision: 0.9374736347553044
DNN Recall: 0.9422338037813368
DNN F1 Score: 0.9316176924270374
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-defined ar
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Recall is ill-defined and b
  _warn_prf(average, modifier, msg_start, len(result))
```

```
#chuyển đổi dữ liệu từ DataFrame sang NumPy arrays
X_np = X.values
y_np = y.values
```

```
# chuyển đổi các giá trị trong mảng NumPy y_np từ chuỗi thành số nguyên
y_np = np.where(y_np == 'BENIGN', 0, 1)
```

```
#sử dụng Min-Max Scaling để chuẩn hóa dữ liệu trong DataFrame df
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(df)
```

```
#bộ dữ liệu huấn luyện (X_tr, y_tr) và bộ dữ liệu kiểm thử (X_t, y_t).
X_tr, X_t, y_tr, y_t = train_test_split(data_scaled[:, :-1], data_scaled[:, -1], test_size=0.2, random_state=42)
```

```
#chuyển đổi mảng hai chiều thành mảng ba chiều để phù hợp với đầu vào mong đợi của một số mô hình học máy; 83 đặc
X_tr = X_tr.reshape((X_tr.shape[0], 1, 83))
X_t = X_t.reshape((X_t.shape[0], 1, 83))
```

```
#lưu trữ thông tin về quá trình huấn luyện mô hình
history = History()
```

```
#xây dựng, biên dịch và huấn luyện một mô hình LSTM
model_lstm = Sequential()
model_lstm.add(LSTM(units=10, input_shape=(X_tr.shape[1], X_tr.shape[2])))
model_lstm.add(Dense(units=1, activation='sigmoid'))
custom_optimizer = Adam(learning_rate=0.0001)
model_lstm.compile(loss='binary_crossentropy', optimizer=custom_optimizer, metrics=['accuracy'])
model_lstm.fit(X_tr, y_tr, epochs=100, batch_size=32, validation_data=(X_t, y_t), callbacks=[history])
```

```
Epoch 81/100
6572/6572 [==============================] - 19s 3ms/step - loss: 0.0010 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy: 0.
Epoch 82/100
6572/6572 [==============================] - 20s 3ms/step - loss: 0.0010 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy: 0.
Epoch 83/100
6572/6572 [==============================] - 19s 3ms/step - loss: 0.0010 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy: 0.
Epoch 84/100
6572/6572 [==============================] - 20s 3ms/step - loss: 0.0010 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy: 0.
Epoch 85/100
6572/6572 [==============================] - 20s 3ms/step - loss: 0.0010 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy: 0.
Epoch 86/100
6572/6572 [==============================] - 19s 3ms/step - loss: 0.0010 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy: 0.
Epoch 87/100
6572/6572 [==============================] - 21s 3ms/step - loss: 9.9874e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 88/100
6572/6572 [==============================] - 20s 3ms/step - loss: 9.9868e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 89/100
6572/6572 [==============================] - 21s 3ms/step - loss: 9.9689e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 90/100
6572/6572 [==============================] - 20s 3ms/step - loss: 9.9710e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 91/100
6572/6572 [==============================] - 20s 3ms/step - loss: 9.9401e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 92/100
6572/6572 [==============================] - 19s 3ms/step - loss: 9.9483e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 93/100
6572/6572 [==============================] - 21s 3ms/step - loss: 9.9228e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 94/100
6572/6572 [==============================] - 19s 3ms/step - loss: 9.9050e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 95/100
6572/6572 [==============================] - 22s 3ms/step - loss: 9.8948e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 96/100
6572/6572 [==============================] - 20s 3ms/step - loss: 9.8841e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 97/100
6572/6572 [==============================] - 20s 3ms/step - loss: 9.8773e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 98/100
6572/6572 [==============================] - 19s 3ms/step - loss: 9.8456e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 99/100
6572/6572 [==============================] - 22s 3ms/step - loss: 9.8608e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
Epoch 100/100
6572/6572 [==============================] - 19s 3ms/step - loss: 9.8239e-04 - accuracy: 0.9999 - val_loss: 0.0016 - val_accuracy
<keras.src.callbacks.History at 0x7c0148413e20>
```

```
#thực hiện dự đoán trên dữ liệu kiểm thử (X_t) bằng mô hình LSTM đã được huấn luyện (model_lstm)
y_pred_prob = model_lstm.predict(X_t)
y_pred = (y_pred_prob > 0.5).astype(int)
```

```
1643/1643 [==============================] - 3s 1ms/step
```

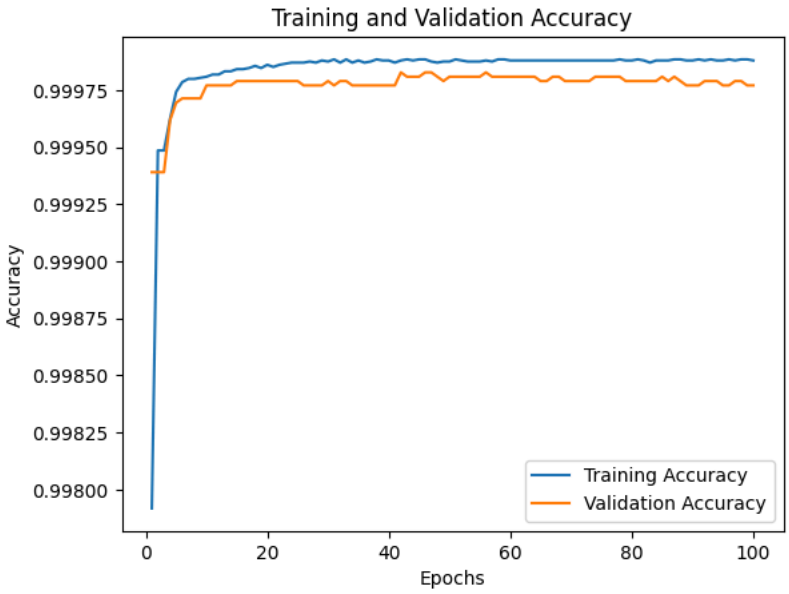
```
#đánh giá hiệu suất của mô hình
precision = precision_score(y_t, y_pred, average='weighted')
recall = recall_score(y_t, y_pred, average='weighted')
accuracy = accuracy_score(y_t, y_pred)
f1 = f1_score(y_t, y_pred, average='weighted')
print(f'Precision: {precision}')
print(f'Recall: {recall}')
print(f'F1 Score: {f1}')
print(f'A: {accuracy}')
```

```
Precision: 0.9997586015889902
Recall: 0.9997717502948226
F1 Score: 0.9997599507914243
A: 0.9997717502948226
```

```
#truy xuất thông tin về độ chính xác (accuracy) của mô hình trên tập huấn luyện và tập kiểm thử qua các epoch tro
training_acc = history.history['accuracy']
validation_acc = history.history['val_accuracy']
```

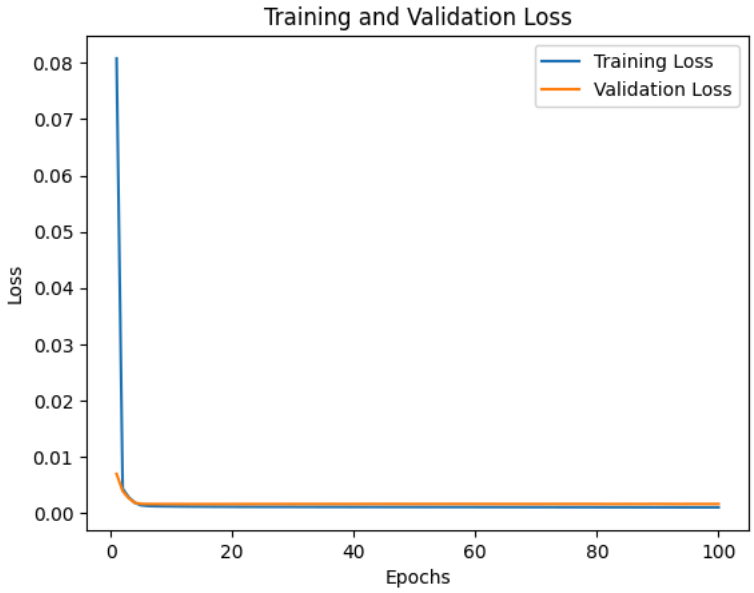
```
import matplotlib.pyplot as plt
# Lấy thông tin về accuracy từ lịch sử đào tạo
train_accuracy = history.history['accuracy']
validation_accuracy = history.history['val_accuracy']
# Tạo mảng với số lượng epochs
epochs = range(1, len(train_accuracy) + 1)
```

```
# Vẽ đồ thị
plt.plot(epochs, train_accuracy, label='Training Accuracy')
plt.plot(epochs, validation_accuracy, label='Validation Accuracy')
plt.title('Training and Validation Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.show()
```




```
# Lấy thông tin về loss từ lịch sử đào tạo
train_loss = history.history['loss']
validation_loss = history.history['val_loss']
# Tạo mảng với số lượng epochs
epochs = range(1, len(train_loss) + 1)
```

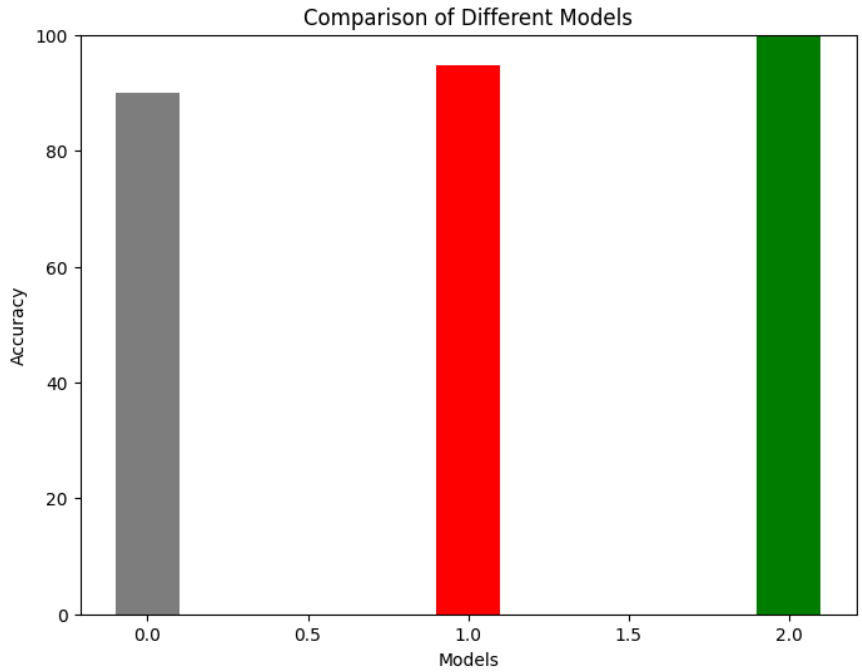
```
# Vẽ đồ thị
plt.plot(epochs, train_loss, label='Training Loss')
plt.plot(epochs, validation_loss, label='Validation Loss')
plt.title('Training and Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()
```




```
import matplotlib.pyplot as plt
import numpy as np
```

```
# Giả sử bạn đã có các mô hình đã được huấn luyện
# Thay thế chúng bằng các mô hình thực tế của bạn
knn_accuracy = 90.08445239091566
dnn_accuracy = 94.75976718530071
lstm_accuracy = 99.97830387132076
models = ['KNN', 'DNN', 'LSTM']
accuracies = [knn_accuracy, dnn_accuracy, lstm_accuracy]
```

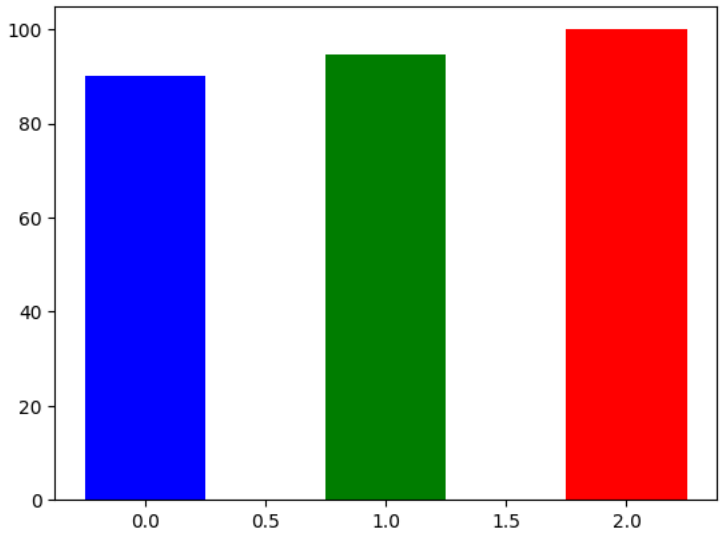
```
# Tạo biểu đồ
plt.figure(figsize=(8, 6))
bar_width = 0.2
positions = np.arange(len(models))
plt.bar(positions, accuracies, width=bar_width, color=['gray', 'red', 'green'], align='center')
# Thêm chú thích cho biểu đồ
plt.title('Comparison of Different Models')
plt.xlabel('Models')
plt.ylabel('Accuracy')
plt.ylim([0, 100])
plt.show()
```




```
import matplotlib.pyplot as plt
import numpy as np
# Giả sử bạn đã có các mô hình đã được huấn luyện
# Thay thế chúng bằng các mô hình thực tế của bạn
knn_accuracy = 90.08445239091566
ann_accuracy = 94.75976718530071
lstm_accuracy = 99.97830387132076
models = ['KNN', 'DNN', 'LSTM']
accuracies = [knn_accuracy, ann_accuracy, lstm_accuracy]


# Màu sắc tương ứng với từng mô hình
colors = ['gray', 'red', 'green']


# Tạo biểu đồ
plt.figure(figsize=(8, 6))
bar_width = 0.2 # Tăng chiều rộng cột
positions = np.arange(len(models))
```

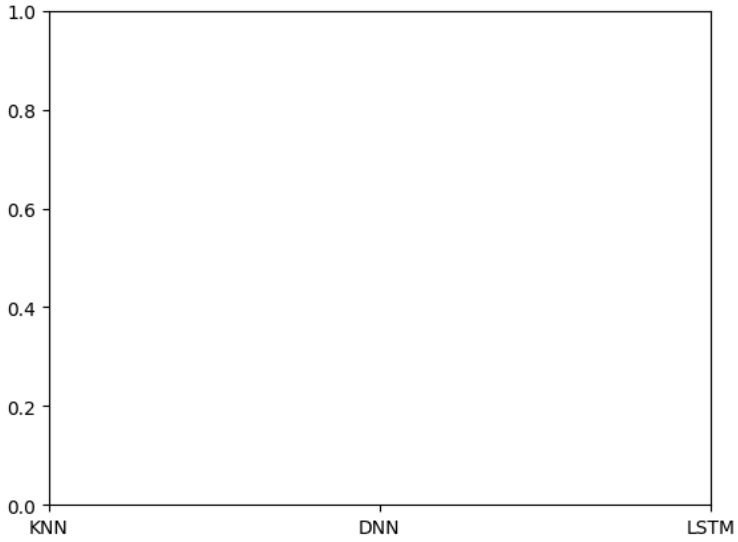
```
<Figure size 800x600 with 0 Axes>
```

```
# Define colors for the bars
colors = ['blue', 'green', 'red']


# Define positions for the bars
positions = np.arange(len(models))


# Define the width of the bars
bar_width = 0.5


import matplotlib.pyplot as plt

# Vẽ biểu đồ với màu sắc tương ứng và nhãn chú thích
for i, (model, accuracy, color) in enumerate(zip(models, accuracies, colors)):
    plt.bar(positions[i], accuracy, width=bar_width, color=color, align='center')

plt.show()
```



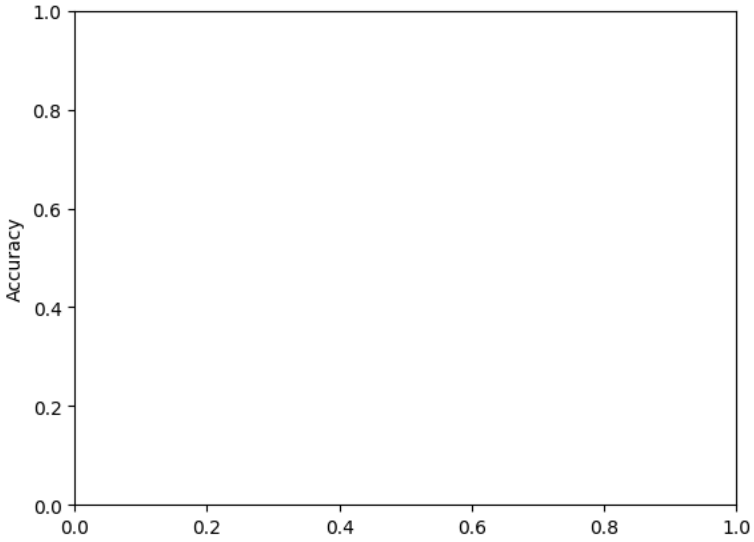

```
# Đặt nhãn cho trục x
plt.xticks(positions, models)
```

```
([<matplotlib.axis.XTick at 0x7c0146071510>,
  <matplotlib.axis.XTick at 0x7c0146071720>,
  <matplotlib.axis.XTick at 0x7c0146071570>],
 [Text(0, 0, 'KNN'), Text(1, 0, 'DNN'), Text(2, 0, 'LSTM')])
```




```
# Đặt nhãn cho trục y
plt.ylabel('Accuracy')
```

```
Text(0, 0.5, 'Accuracy')
```



```
# Hiển thị biểu đồ
plt.show()
```