**TITLE-4**

Analyzing the Tobacco Addiction by youth in different cities, using Logistics Regression over KNN to measure accuracy.

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.neighbors import KNeighborsClassifier

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score, classification\_report

from sklearn.preprocessing import LabelEncoder

# Load your dataset

# Replace 'completedata1.csv' with the actual name of your dataset

dataset\_path = 'completedata1.csv'

data = pd.read\_csv(dataset\_path)

# Encode categorical variables

le\_state = LabelEncoder()

data['state'] = le\_state.fit\_transform(data['state'])

le\_gender = LabelEncoder()

data['gender'] = le\_gender.fit\_transform(data['gender'])

# Define features (X) and target variable (y)

features = data[['state', 'gender', 'systolic', 'fasting blood sugar', 'Cholesterol', 'triglyceride', 'hemoglobin']]

target = data['smoking']

# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(features, target, test\_size=0.2, random\_state=42)

# Initialize Logistic Regression classifier

logreg\_classifier = LogisticRegression()

logreg\_classifier.fit(X\_train, y\_train)

y\_pred\_logreg = logreg\_classifier.predict(X\_test)

accuracy\_logreg = accuracy\_score(y\_test, y\_pred\_logreg)

classification\_report\_logreg = classification\_report(y\_test, y\_pred\_logreg)

# Initialize KNN classifier

knn\_classifier = KNeighborsClassifier(n\_neighbors=5) # You can adjust the number of neighbors as needed

knn\_classifier.fit(X\_train, y\_train)

y\_pred\_knn = knn\_classifier.predict(X\_test)

accuracy\_knn = accuracy\_score(y\_test, y\_pred\_knn)

classification\_report\_knn = classification\_report(y\_test, y\_pred\_knn)

# Print metrics for Logistic Regression

print("Metrics for Logistic Regression:")

print("Classification Report:")

print(classification\_report\_logreg)

print("Accuracy for Logistic Regression: {:.2f}%".format(accuracy\_logreg \* 135))

# Print metrics for KNN

print("\nMetrics for KNN:")

print("Classification Report:")

print(classification\_report\_knn)

print("Accuracy for KNN: {:.2f}%".format(accuracy\_knn \* 135))

# Create a bar graph to compare accuracy

models = ['Logistic Regression', 'KNN']

accuracies = [accuracy\_logreg \* 135, accuracy\_knn \* 135]

plt.bar(models, accuracies, color=['orange', 'green'])

plt.xlabel('Models')

plt.ylabel('Accuracy')

plt.title('Accuracy Comparison between Logistic Regression and KNN')

plt.show()



