Text

Description automatically generated

**Activity based**

**Project Report on**

**DWDM**

**Submitted to Vishwakarma University, Pune**

**Under the Initiative of**

**By**

**Arun Adhikari**

**SRN No : 202100406**

**Roll No : 38**

**Div : G**

**Third Year Engineering**

**Department of Computer Engineering**

**Faculty of Science and Technology**

**Academic Year**

**2023-2024**

**Problem Statement**

**Implement a k-means Clustering Algorithm.**

**Business Requirements**

**Stakeholders: Executive leadership, customers,**

**Data Sources: Customer database, content database**

import pandas as pd

from scipy.cluster.hierarchy import dendrogram, linkage, fcluster

from sklearn.preprocessing import StandardScaler

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

file\_path = 'vehicle telemetry.csv' # Update this with the path to your CSV file

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

# Data Preparation: Handle missing values in 'horsepower'

data['horsepower'].fillna(data['horsepower'].median(), inplace=True)

# Select features for clustering (e.g., cylinders, displacement, horsepower, weight, acceleration)

X = data[['cylinders', 'displacement', 'horsepower', 'weight', 'acceleration']]

# Normalize the features

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

# Perform hierarchical clustering

# 'ward' minimizes the variance of the clusters being merged

Z = linkage(X\_scaled, method='ward')

# Plot the dendrogram

plt.figure(figsize=(12, 8))

dendrogram(Z, truncate\_mode='level', p=5, labels=data.index, leaf\_rotation=90, leaf\_font\_size=10)

plt.title("Dendrogram for Hierarchical Clustering")

plt.xlabel("Data Point Index")

plt.ylabel("Euclidean Distance")

plt.show()

# Choose the number of clusters by cutting the dendrogram at a specific distance

# For example, setting a threshold that corresponds to 3 clusters

num\_clusters = 3

data['cluster'] = fcluster(Z, num\_clusters, criterion='maxclust')

# Plot the clusters (e.g., based on displacement and horsepower)

plt.figure(figsize=(10, 6))

sns.scatterplot(x=data['displacement'], y=data['horsepower'], hue=data['cluster'], palette='viridis')

plt.title("Hierarchical Clustering of Vehicles")

plt.xlabel("Displacement")

plt.ylabel("Horsepower")

plt.legend(title="Cluster")

plt.show()



