

Data Mining Final Practical File

24/48029

Vivaan Singh Adhikari November 24, 2025

Assignment Overview

Document Contents:

1. Introduction	2
1.1. Code	2

Assignment Details:

- **Course:** Data Mining DSE
- **Instructor:** Prof. Archana Gahalaut
- **Hardware:** No specifications
- **Software:** Python, Pandas, Typst(documentation)

Introduction

This assignment entails my solutions to the question assigned as per the course's guidelines. All the final files are available on <https://github.com/user7537/coursework/>

Code

QApply simple K-means algorithm for clustering any dataset. Compare the performance of clusters by varying the algorithm parameters. For a given set of parameters, plot a line graph depicting MSE obtained after each iteration.

```
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules

df1 = pd.read_csv("japanese_credit_screening/crx.data", header=None,
na_values="?")
df2 = pd.read_csv("db2/SouthGermanCredit.asc", sep=" ", header=None)

df1.columns = [f"A{i}" for i in range(1,17)]
df2.columns =
["laufkont", "laufzeit", "moral", "verw", "hoehe", "sparkont", "beszeit", "rate", "famges", "buerge", "wohnm"]

df1 = df1.fillna("MISSING")

for c in df1.columns:
    df1[c] = df1[c].astype(str)
for c in df2.columns:
    df2[c] = df2[c].astype(str)

df1_onehot = pd.get_dummies(df1)
df2_onehot = pd.get_dummies(df2)

freq1_A = apriori(df1_onehot, min_support=0.5, use_colnames=True)
rules1_A = association_rules(freq1_A, metric="confidence",
min_threshold=0.75)
freq1_A.to_csv("q3_ds1_freq_50_75.csv", index=False)
rules1_A.to_csv("q3_ds1_rules_50_75.csv", index=False)

freq1_B = apriori(df1_onehot, min_support=0.6, use_colnames=True)
rules1_B = association_rules(freq1_B, metric="confidence",
min_threshold=0.60)
freq1_B.to_csv("q3_ds1_freq_60_60.csv", index=False)
rules1_B.to_csv("q3_ds1_rules_60_60.csv", index=False)
```

```
freq2_A = apriori(df2_onehot, min_support=0.5, use_colnames=True)
rules2_A = association_rules(freq2_A, metric="confidence",
min_threshold=0.75)
freq2_A.to_csv("q3_ds2_freq_50_75.csv", index=False)
rules2_A.to_csv("q3_ds2_rules_50_75.csv", index=False)

freq2_B = apriori(df2_onehot, min_support=0.6, use_colnames=True)
rules2_B = association_rules(freq2_B, metric="confidence",
min_threshold=0.60)
freq2_B.to_csv("q3_ds2_freq_60_60.csv", index=False)
rules2_B.to_csv("q3_ds2_rules_60_60.csv", index=False)

print("done")
```