TITLE

Project Report

SUPERVISOR: Supervisor Name

SUBMITTED BY

Name1 Exam Roll Number1

Name2 Exam Roll Number2

2022

Department of Computer Science

ACHARYA NARENDRA DEV COLLEGE

ACKNOWLEDGEMENT

General Instructions for the report

1. Running Text should be in 12 pt Times New Roman Justified on both sides.
2. Captions of tables should be on top of table while of figures, captions should be below figures. Keep caption of tables/figures in 10 Pt. Times New Roman.
3. There should not be any free space left after tables or figures at the bottom of the page. Adjust your text, figures and tables such that it should look a continuous document.
4. Chapter Title : Times New Roman 16 Pt Bold center Aligned
5. Head1 Title:... Times New Roman 14 Pt. Bold
6. Head2 Title….Times New Roman 12 Pt. Bold
7. Margins…Normal: 1 inch from top, bottom, left, right
8. Inter line spacing: 1.15..... ....

....

name 1

name 2

ACHARYA NARENDRA DEV COLLEGE

(University of Delhi)

CERTIFICATE

Certificate text. . . . .
.
. . ..
. . .
. . ..
. .
.. . .
. . .
. .. .
. . .
. . ..
. . .
. . ..
. .
.. . .
. . .
. .. .
. . .
. . ..

Supervisor
Supervisor name

Table of Contents

Chapter 1

PROBLEM STATEMENT

Describe in detail what problem you are solving and why.

Chapter 2

DATA MINING TECHNIQUES

2.1. DM Techniques

Describe briefly different DM techniques.

2.1.1. Classification

2.1.2. Association

2.1.3. Clustering

2.2. **Data mining technique used for this project.
say Classification **

2.2.1. K-NN

2.2.2. Naive Bayes

2.2.3. Decision Tree

Chapter 3

Dataset Description

3.1 Dataset

Briefly describe the dataset and its source.

for each of the following use appropriate Python/R methods and paste output screenshots

3.1.1. Number of Records
3.1.2 Number of Attributes
3.1.3. Types of Attributes
3.1.4. Missing Values or Nulls
3.1.5. Attributes Description
3.1.6. Distribution/Histograms
3.1.7. Detecting Outliers

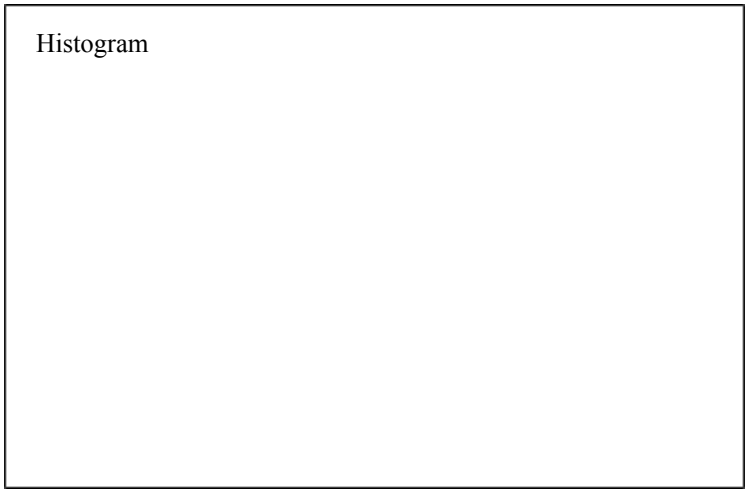

Figure 3.1: Histogram

Figure 3.2: Outliers

Chapter 4

DATA PREPROCESSING

Briefly describe importance of data preprocessing and what all comprises of data preprocessing

4.1 HANDLING NULL VALUES

Talk about why handling null/NaN values is important. What are different reasons for NULL values. What are the different ways to handle NULL.

Demonstrate various ways you have handled NULLs in your Project. (Share commands/functions and screenshots)

4.2 FEATURE SCALING

Describe feature scaling and describe

4.2.1 Normalization

4.2.2 Standardization

4.2.3 Demonstrate Feature Scaling in your project.

4.3 FEATURE SELECTION AND CONVERSION

What is Feature selection? What features are important for your problem statement?

What are the various attributes that require conversion for addressing your project problem?

4.3.1 CATEGORICAL TO NUMERICAL

4.3.2 . . . . . .

4.3.3 . . . . . .

4.4 DATA SAMPLING AND SUBSETTING

Why do we need to split the data?

4.4.1 What are various ways of sampling data?

4.4.2 TRAIN-TEST SPLIT

Demonstrate through Python/R code and display train/test subsets

Chapter 5

Building Models

Describe why you need to train and test different models?

5.1 Model1

Describe model: Features used etc

5.2 Model2

Describe model: Features used etc

5.3 Model3

Describe model: Features used

Chapter 6

MODEL EVALUATION AND RESULTS

6.1 METRICS

How do we compare models? What are the various metrics used. Describe them briefly.

- 6.1.1 CONFUSION MATRIX
- 6.1.2 ACCURACY
- 6.1.3 PRECISION
- 6.1.4 RECALL
- 6.1.5 F1-SCORE

6.2. EXPERIMENTAL RESULTS AND COMPARISON

Report CM, ACC, Prec, Recall, F1 for all models discussed in Chapter 5

Chapter 7

INFERENCES AND CONCLUSION

Discuss inferences.

Based on the Experimental Results reported in Chapter 6 which models will you recommend and why.

Reference

Please give all references in APA style. Cite all books, data resources, reading material, research papers, journal articles, etc. referred to in APA style.

1. Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
2. Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
3. Kesavaraj, G., & Sukumaran, S. (2013, July). A study on classification techniques in data mining. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)* (pp. 1-7). IEEE.