
**PUNE INSTITUTE OF COMPUTER TECHNOLOGY,**

**PUNE- 411043**

**Department of Computer Engineering**

**S.No.-27, Pune Satara Road,**

**Dhankawadi, Pune-411043**

**Laboratory Practice VI**

**Class: BE**

**Batch: P3**

**(2022-2023)**

# Business Intelligence Mini Project Report

**Group members (Roll no and Name):**

| Roll No. | Name |
|---|---|
| 41303 | Rutuja Banginwar |
| 41305 | Vrushaket Chaudhari |

**Lab Teacher Name:** Prof. Pujashree Vidap

## Title:

A mini project on: BI Report for data mining task.

## Problem Statement:

A BI report must be prepared outlining the following steps:

a) Problem definition, identifying which data mining task is needed.

b) Identify and use a standard data mining dataset available for the problem.

## Objectives and Scope:

- Objective of this project is to understand BI techniques and be able to prepare Report on it.
- Learned how to perform data mining.

## Software and Hardware Requirements:

- Processor: Intel i5 gen
- RAM: 8GB
- Operating System: Windows 10
- Jupyter notebook

# Business Intelligence Report

### 1. Problem Definition and Data Mining Task

**Objective:** The objective of this report is to classify Iris flowers into different species based on their features using the K-Nearest Neighbors algorithm.

**Data Mining Task:** Classification

### 2. Standard Data Mining Dataset Selection

**Dataset:** Iris Dataset

**Source:** The Iris dataset is a standard dataset widely used in the field of data mining. It was originally introduced by Ronald Fisher in 1936 and is now commonly available in various repositories and libraries.

### 3. Dataset Overview and Understanding

The Iris dataset provides the following information:

- **Structure:** The dataset contains a total of 150 records, with each record representing an Iris flower. There are four attributes or features available for each flower, namely sepal length, sepal width, petal length, and petal width. The dataset is evenly distributed among three different species of Iris flowers, with 50 records for each species.

- **Variables:**
  - Sepal Length (in cm): The length of the sepal of an Iris flower.
  - Sepal Width (in cm): The width of the sepal of an Iris flower.
  - Petal Length (in cm): The length of the petal of an Iris flower.
  - Petal Width (in cm): The width of the petal of an Iris flower.
  - Species: The species of the Iris flower, which can be one of the following three classes: Setosa, Versicolor, and Virginica.

### 4. Data Preparation

To prepare the dataset for the K-Nearest Neighbors classification analysis, the following preprocessing steps were performed:

- **Data Cleaning:** The dataset was checked for missing values, and no missing values were found.

- **Data Transformation:** No specific data transformations were applied to the dataset.

- **Data Encoding:** The categorical variable "Species" was encoded into numerical values for the classification analysis. Setosa=0, Versicolor=1, Virginica=2.

### 5. Data Mining Process

For the classification analysis of the Iris dataset, the K-Nearest Neighbors algorithm was

employed:

- **Data Mining Technique:** K-Nearest Neighbors (KNN)

- **Model Training and Evaluation:** The KNN classifier was trained on 70% of the dataset using a stratified train-test split. The number of neighbors (K) was set to [choose an appropriate value]. The model's performance was evaluated using accuracy, precision, recall, and F1-score metrics.

**6. Analysis and Interpretation of Results**

The classification analysis of the Iris dataset using the K-Nearest Neighbors algorithm yielded the following key findings:

- The KNN classifier achieved an accuracy of **1.0** on the test dataset, indicating its effectiveness in classifying Iris flowers into different species.
- The optimal number of neighbors (K) was determined to be **3** based on the performance metrics and cross-validation analysis.
- The KNN model successfully classified the Iris flowers into their respective species with high precision, recall, and F1-scores.

**7. Conclusion and Recommendations**

Based on the analysis conducted, the following conclusions and recommendations are provided:

- **Conclusion:** The K-Nearest Neighbors algorithm demonstrated its suitability for the classification of Iris flowers based on their features. The model achieved high accuracy and performed well in terms of precision, recall, and F1-scores.
- **Recommendations:** It is recommended to utilize the trained KNN model to classify new Iris flowers based on their features. Additionally, further exploration can be done by comparing the performance of other classification algorithms on the Iris dataset to identify the most suitable model for specific business requirements.

## Python Implementation:

```
# Import the required libraries
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, classification_report

# Load the Iris dataset
iris = load_iris()
X = iris.data
y = iris.target

# Split the dataset into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)

# Create a K-Nearest Neighbors classifier object
knn = KNeighborsClassifier(n_neighbors=3)  # Specify the number of
neighbors (K)

# Train the classifier using the training data
knn.fit(X_train, y_train)

# Make predictions on the test data
y_pred = knn.predict(X_test)

# Evaluate the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Generate a classification report
report = classification_report(y_test, y_pred,
target_names=iris.target_names)
print("Classification Report:\n", report)
```

## Output:

```
Accuracy: 1.0
Classification Report:
              precision    recall  f1-score   support

      setosa       1.00      1.00      1.00        19
  versicolor       1.00      1.00      1.00        13
   virginica       1.00      1.00      1.00        13

    accuracy                           1.00        45
   macro avg       1.00      1.00      1.00        45
weighted avg       1.00      1.00      1.00        45
```

**Conclusion:**

Thus we successfully prepared a  BI Report and performed the data mining task.