PROJECT REPORT

Preparing a report for Iris prediction data involves several steps, from understanding the dataset to presenting the results of analysis .

1. Title Page

Title: Clear and concise title (e.g., "Iris Flower Species Prediction Report").

Your Name: Prince and Mohit Kumar

Date: 21/02/2025

Purpose: To analyse the Iris dataset and predict species using machine learning models

2. Introduction

Background: It is a classic dataset used for classification tasks, containing measurements of Iris flowers**.**

Objective: To predict Iris species based on sepal and petal measurements using machine learning algorithms

Dataset Overview: 3 classes: Setosa, Versicolor, Virginica**.**

3. Dataset Description

Features: Describe the features in the dataset:

Sepal Length (cm)

Sepal Width (cm)

Petal Length (cm)

Petal Width (cm)

Target Variable: Iris species: Setosa, Versicolor, Virginica.

Data Source: Mention the source of the dataset (Kaggle).

Exploratory Data Analysis (EDA):

Include visualizations (histograms) to show the distribution of features and relationships between them.

Using correlation methods with seaborn library for better visualizations.

4. Methodology

Data Preprocessing:

handling missing values, scaling, encoding categorical variables .

Train-Test Split: Explain how the data was split into training and testing sets (e.g., 80% training, 20% testing).

Model Selection: Logistic Regression, Decision Trees, K-Nearest Neighbors

Evaluation Metrics: Specify the metrics used to evaluate the models (e.g., accuracy, precision, recall, F1-score, confusion matrix).

5. Model Training and Evaluation

Model Training: Briefly describe how each model was trained.

Results:

Present the performance of each model using tables or charts.

Compare the models based on the evaluation metrics.

Best Model: Use Logistic Regression with the accuracy of 98%.

7. Conclusion

Summary: Summarize the key findings and outcomes of the analysis.

Recommendations: Suggest next steps (e.g., trying more advanced models, collecting more data, or applying the model to a real-world scenario).

7. Conclusion

Summary of Key Findings

Model Performance: Recap the performance of the models .

"The K-Nearest Neighbors (KNN) model achieved the highest accuracy of 98%, outperforming Logistic Regression and Decision Trees."

"The confusion matrix showed that the model performed exceptionally well in classifying Setosa and Versicolor but had slight difficulty distinguishing between Versicolor and Virginica."

Feature Importance: Highlight any insights about the features:

"Petal length and petal width were the most significant features for predicting Iris species, as observed in the feature importance analysis."