**Predictor Comparison Project**

This project compares the performance of **Logistic Regression** and **Support Vector Machines (SVM)** using the **Wine dataset** from sklearn.datasets. The comparison involves generating **accuracy metrics**, **decision boundary visualizations**, and **bar/line charts** for both models across various hyperparameter values. The results are compiled into a Word document report.

**Project Structure**

- generate\_report.py (Main script to generate the report)

- report/ (Folder where the report and plots will be saved)

- README.docx (This README file)

**Installation**

**Prerequisites**

1. **Python 3.8 or higher**
2. **Virtual environment** (optional, I used conda environment in Visual Studio Code)

**Python Libraries**

The required Python libraries are:

* matplotlib
* numpy
* scikit-learn
* python-docx

You can install the dependencies using pip:

pip install matplotlib numpy scikit-learn python-docx

**Running the Project**

To generate the report, follow these steps:

1. Clone the repository or download the generate\_report.py script.
2. Navigate to the project directory.
3. Run the generate\_report.py script:

python generate\_report.py

**The script will:**

* Train **Logistic Regression** and **SVM** models on the Wine dataset.
* Calculate accuracies and generate visualizations for different **C** values (Logistic Regression) and **gamma** values (SVM).
* Save all plots (decision boundaries, bar charts, line charts) in the report/ folder.
* Generate a **Word report** (Predictor\_Comparison\_Report.docx) with charts and my comments summarizing the results.

**Project Workflow**

1. **Data Loading & Preprocessing**:
   * The Wine dataset is loaded from sklearn.datasets.
   * The features "Alcohol" and "Malic Acid" are selected and standardized.
2. **Model Training**:
   * **Logistic Regression** is trained with various regularization values (C).
   * **SVM** is trained with different kernel width parameters (gamma).
3. **Visualization**:
   * **Decision boundary plots**: Visualize how each model separates the classes based on different hyperparameter values.
   * **Accuracy bar charts**: Compare the accuracy of each model for different values of C (Logistic Regression) and gamma (SVM).
   * **Accuracy line charts**: Show the accuracy trend across the hyperparameter range.
4. **Report Generation**:
   * A **Word report** (Predictor\_Comparison\_Report.docx) is generated in the report/ folder. It includes all the charts and technical insights.

**Output**

After running the script, the following outputs will be generated in the report/ folder:

1. **Bar Charts**:
   * Logistic Regression Accuracy for All Values
   * SVM Accuracy for All Values
   * Filtered Bar Charts (e.g., excluding specific values).
2. **Line Charts**:
   * Logistic Regression Accuracy Line Chart
   * SVM Accuracy Line Chart
3. **Decision Boundary Plots**:
   * Plots showing how Logistic Regression and SVM classify the Wine dataset for different hyperparameter values.
4. **Report**:
   * Predictor\_Comparison\_Report.docx: A Word document containing all the charts and an analysis of the results.

**License**

This project is licensed under the MIT License. You can find the license terms in the LICENSE file of this project.

**Acknowledgments**

* This project uses the Wine dataset from sklearn.datasets.
* Special thanks to the developers of matplotlib, scikit-learn, and python-docx for providing the tools that made this project possible.