**A purple circle with dots and lines

Description automatically generated**

**NAME:**

**Basit Abbas**

**ROLL NUMBER:**

**SU92-BSDSM-F23- 013**

**PROGRAM:**

**BS DATA SCIENCE**

**SUMBITTED TO:**

**Sir Rasikh**

***SEMESTER PROJECT***

**Project Report: Handwritten Digit Recognition using Machine Learning**

**Introduction:**

Handwritten digit recognition is a fundamental machine learning problem that has wide applications in postal automation, banking systems, and document digitization. This project aims to develop a machine learning model to classify handwritten digits using the Scikit-learn library and the load\_digits dataset. The ultimate objective is to implement a web-based application to demonstrate real-time predictions.

**Objectives:**

1. **Loading Dataset:**  
To load the load\_digits dataset provided by Scikit-learn, which consists of 8x8 grayscale images of handwritten digits. This serves as the foundation for the project by providing the required data for training and testing the model.

2. **Data Exploration**  
To analyze and understand the dataset by visualizing the digit images and examining the distribution of classes. This step helps in identifying any potential challenges, such as class imbalances.

3. **Preprocessing**  
To preprocess the data by normalizing the pixel values to a 0–1 range and ensuring that the dataset is clean and ready for training. This step also includes reshaping or transforming the data as needed for machine learning algorithms.

4. **Splitting**  
To divide the dataset into training and testing subsets, typically in an 80:20 ratio. The training set is used to build the model, while the testing set is reserved for evaluating its performance.

5. **Training/Applying Classifiers**  
To train machine learning classifiers, particularly a Support Vector Machine (SVM), on the training dataset. Experimentation with various kernels and hyperparameters will be conducted to achieve the best performance.

6. **Testing & Processing Results**  
To evaluate the trained model on the test dataset by calculating metrics such as accuracy, precision, recall, and confusion matrix. This step ensures that the model generalizes well to unseen data.

7. **Displaying Accuracies**  
To visually display the accuracy and other evaluation metrics, comparing the performance of different classifiers (if multiple are used). This provides insight into which model performs best for the task.

8. **Application Phase**  
To deploy the trained model in a web-based application using Streamlit. This phase allows users to interact with the model by uploading or drawing handwritten digits to see predictions in real time.

**Tools and Technologies:**

1. **Programming Language**: Python
2. **Libraries and Frameworks**:
   * **Scikit-learn**: For machine learning modeling and evaluation.
   * **Pandas & NumPy**: For data manipulation.
   * **Matplotlib**: For data visualization.
   * **Streamlit**: For creating the web application.
3. **Dataset**: load\_digits from Scikit-learn.
4. **Environment**: VS Code for coding and Streamlit for application development.

**Conclusion**

* The Support Vector Classifier (SVC) model successfully classifies handwritten digits with a high level of accuracy.
* The model can be further improved by experimenting with different kernels or using more advanced techniques like GridSearch for hyperparameter tuning.
* This project demonstrated the basic steps of applying machine learning to an image classification problem, from data preprocessing to training and evaluation.