**Summary**

**Introduction:**

* As of 2024, the UK gambling industry comprises over 2,500 licensed operators.
* For example, X alone operates approximately 1350 betting shops across the UK. 2700 employees work under company X (Approx 2 in a shop)
* 90% of Bets are placed and translated manually.
* To manage this task, X company employs approximately 50% staff members dedicated to translating and placing these bets efficiently.

**Overview of the dataset:**

* Dataset contains a training set of 60,000 images along with labels.
* A test set of 10,000 images of handwritten digits.
* The handwritten digit images have been size normalized and centered in a fixed size of 28×28 pixels.

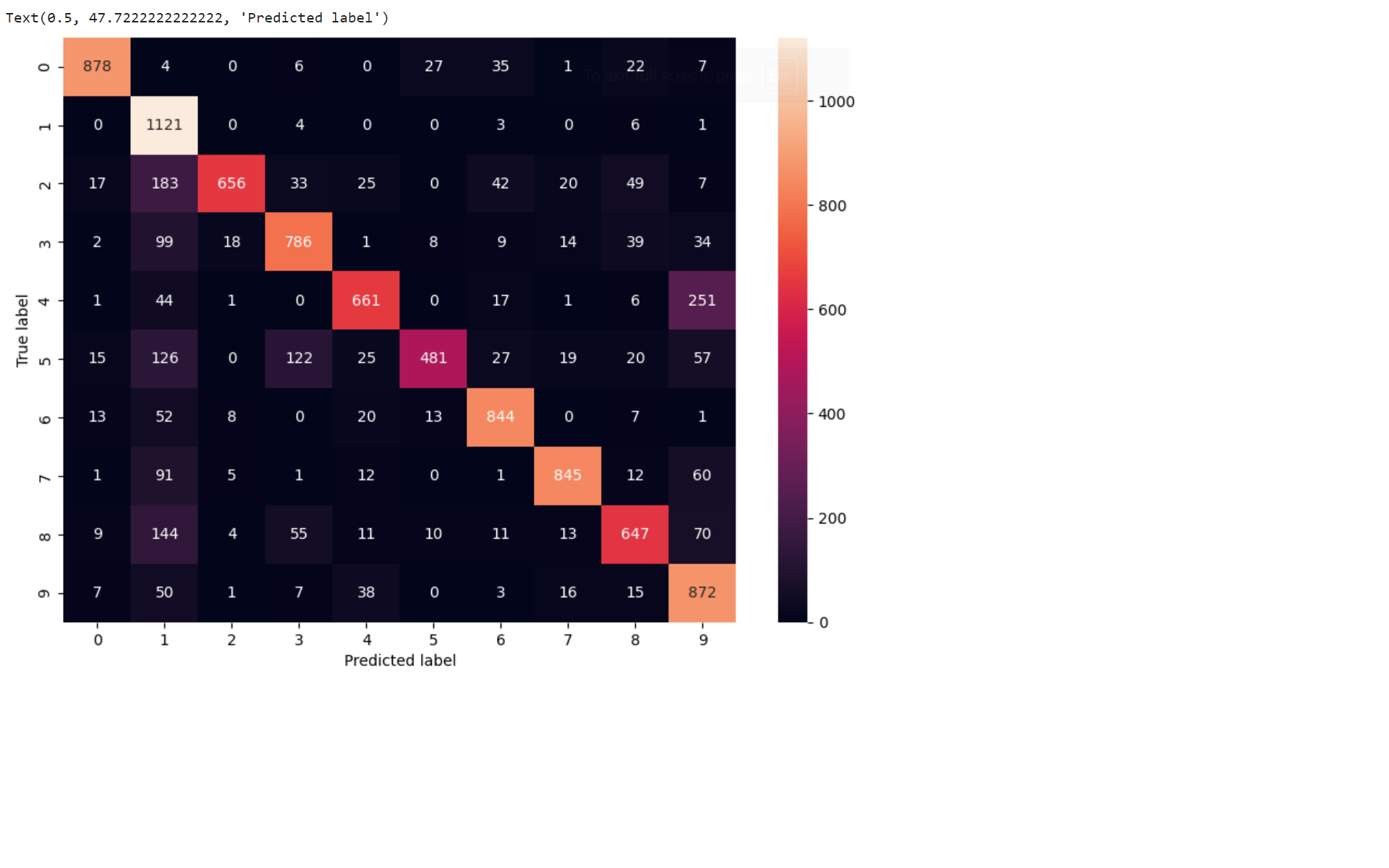
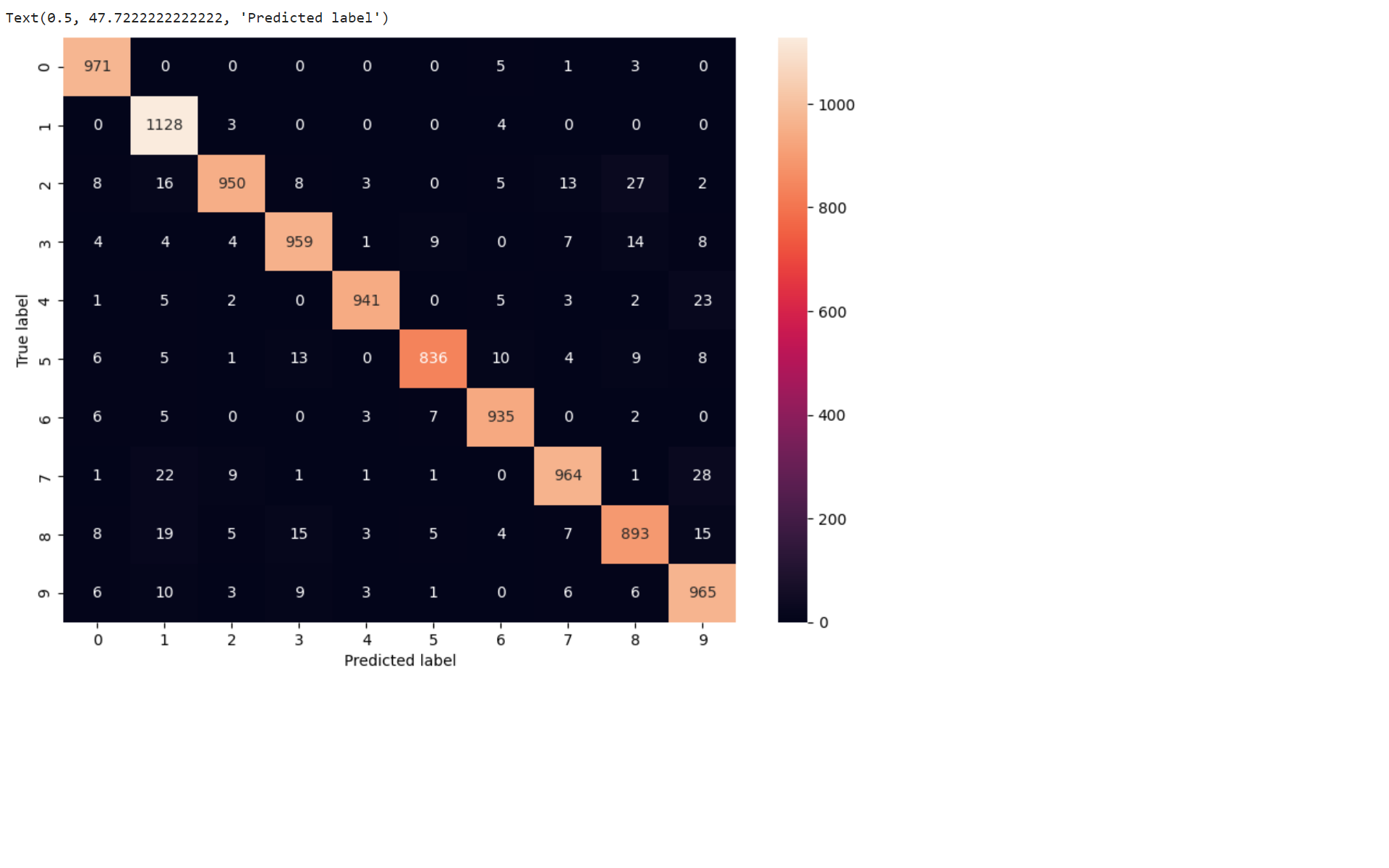
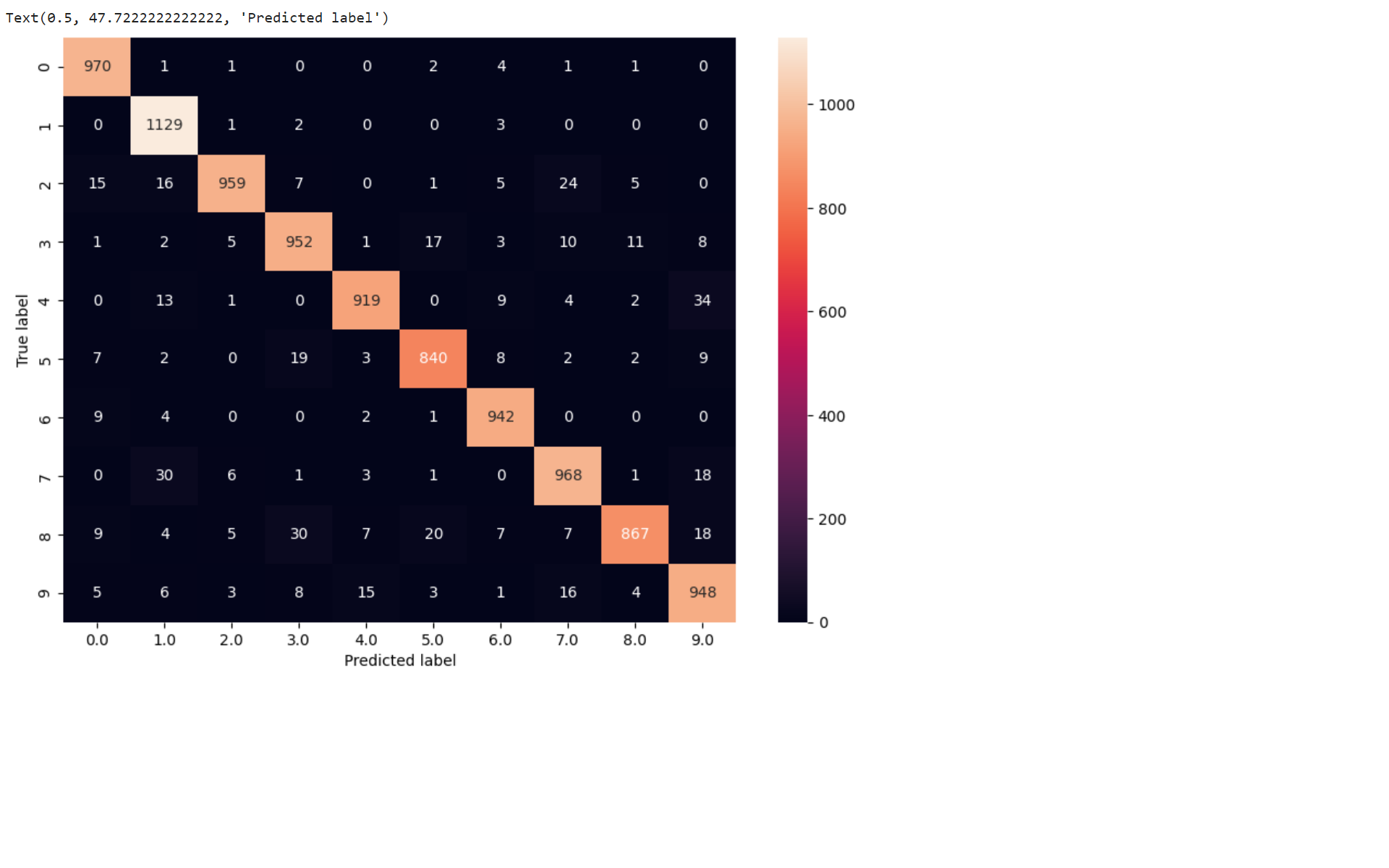
# **Data Preprocessing:**

# **Reshaping:** The images, originally represented as 784-dimensional vectors, were reshaped into 28\*28 matrices for visualization purposes.

# **Normalization:** The pixel values were normalized to the range of 0 to 1 to improve the efficiency of the algorithm. Basically, each value is divided by the highest value i.e 255.

# **Tuning:** The epsilon value taken is 1e-1, which is tuned in such a way that model provides highest accuracy.

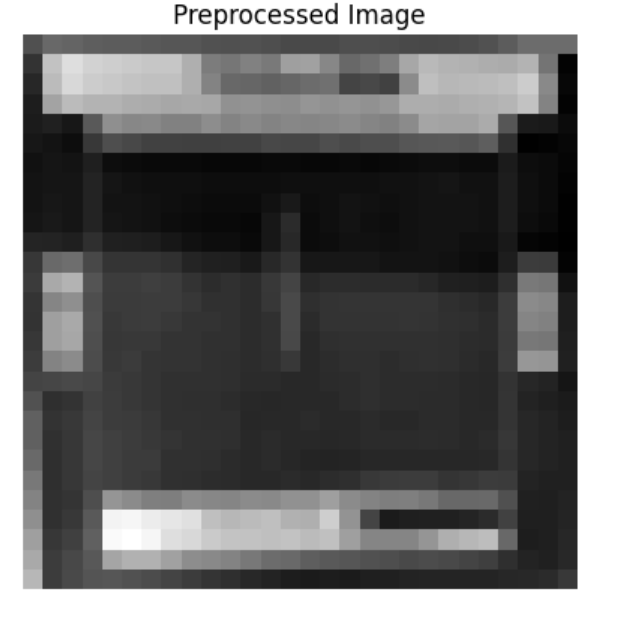
**Implementation:**

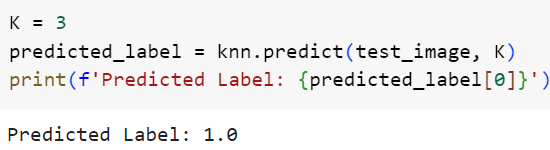
* **Naive Bayes:** A manual implementation of the NB algorithm was used, which calculates the Euclidean distance between a test image and all training images. The algorithm then predicts the label of the test image based on the most frequent label.
* **Accuracy:** 76.51%
* 
* **Gauss Bayes:** A manual implementation of the Gauss algorithm was used, which calculates the Euclidean distance between a test image and all training images. The algorithm then predicts the label of the test image based on the most frequent label.
* **Accuracy:** 95.42%
* 
* **K-Nearest Neighbors (KNN):** A manual implementation of the KNN algorithm was used, which calculates the Euclidean distance between a test image and all training images. The algorithm then predicts the label of the test image based on the most frequent label among the K nearest neighbors.
* **Accuracy:** 94.94%
* 

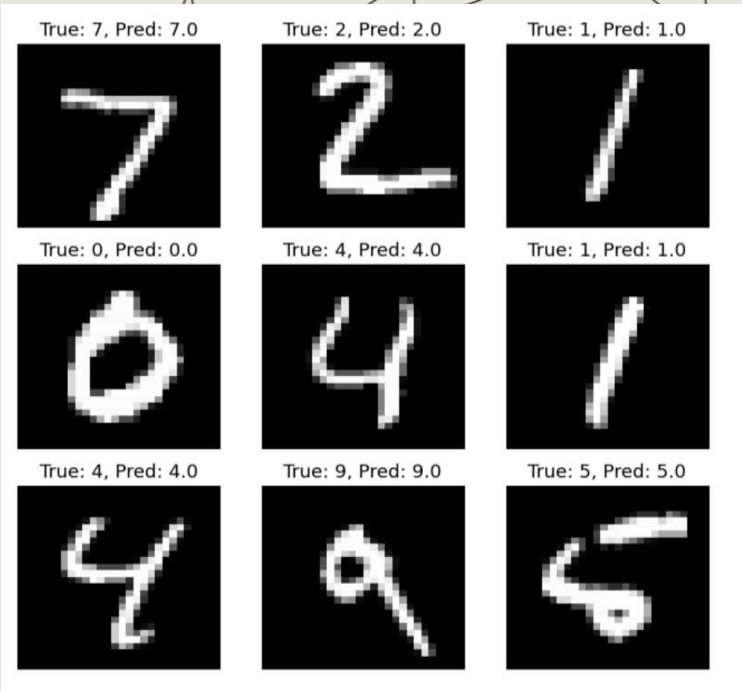
**Tools Used:**

**Python Libraries** used Numpy, Pandas used for Data processing and Normalization process and Matplotlib used for Data visualization, pillow for image processing.

**Conclusion:**







* Gauss Naïve bayes has highest accuracy = 95.42%
* By implementing this system, we can enhance the betting system and reduce costs to the company.
* Further, it helps to be more eco-friendly by reducing paper usage.
* Digital transformation of betting system.