

Tip Recognition

Jacob Alamparambil

Main Problem

SEQ #: 20
Batch #: 288
INVOICE 20
SERVER 0095
Approval Code: 01862B
Entry Method: Chip Read
Mode: Issuer

*# tip the
Bill Challenge*

PRE-TIP AMT \$87.46
TIP 87.46
TOTAL AMOUNT 174.92

*Thank
you
for great
service.*

Overview

- Load csv files
- Divide data into training and testing set
- Normalize pixel values
- Gaussian Non Naive Bayes classifier
- Fit method
- Predict method
- Confusion matrix
- Results
- Conclusion

Step1: Load tip data

```
# Load CSV data
train_df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/MNIST_train.csv")
test_df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/MNIST_test.csv - MNIST_test.csv.csv")
```

Step2: Divide training and testing data from csv file

- Identified features for training set and updated dataframe to only include these columns
- Identified labels for training set
- Repeated this same procedure for testing set

Step 3: Normalize pixel values

- Make sure digits are in black and white for digit recognition

Gaussian Non Naive Bayes Classifier

- Definition
- Fit method
- Predict Method

Step4: Fit training data

1. Identify which samples belong to each class
 - a. Find all training images for our condition
2. Compute class mean vector
 - a. Take average value among all images of that class
3. Compute class covariance matrix
 - a. How pixel intensities vary together within class K
4. Compute class prior
 - a. Probability of randomly seeing class k
5. Store all parameters
 - a. Load up parameters to use in predict function

Step5: Predict

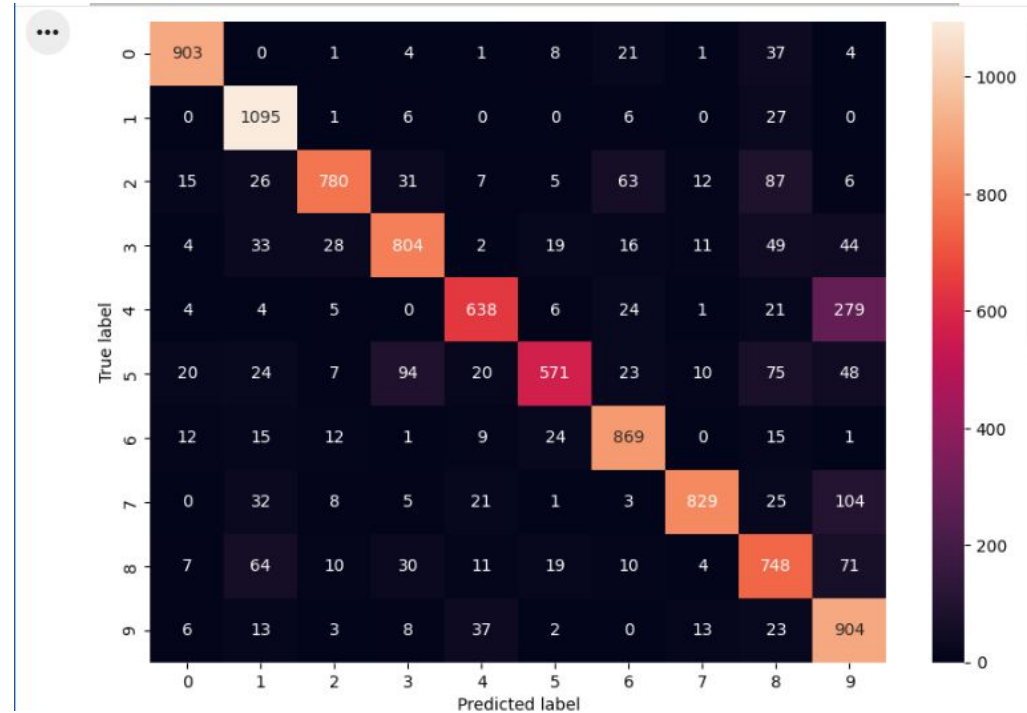
1. Compute how well each digit model matches input image
2. Compute how common digit occurs
3. Combine steps 1 and 2 to compute total score for digital k value
4. Return the highest score

Step 6: Confusion Matrix (Naive Bayes)

This matrix displays the frequency count of each digit in a heatmap format

Min:4 Max:1 Median:804/829->816.5

Model Accuracy: 81.41%



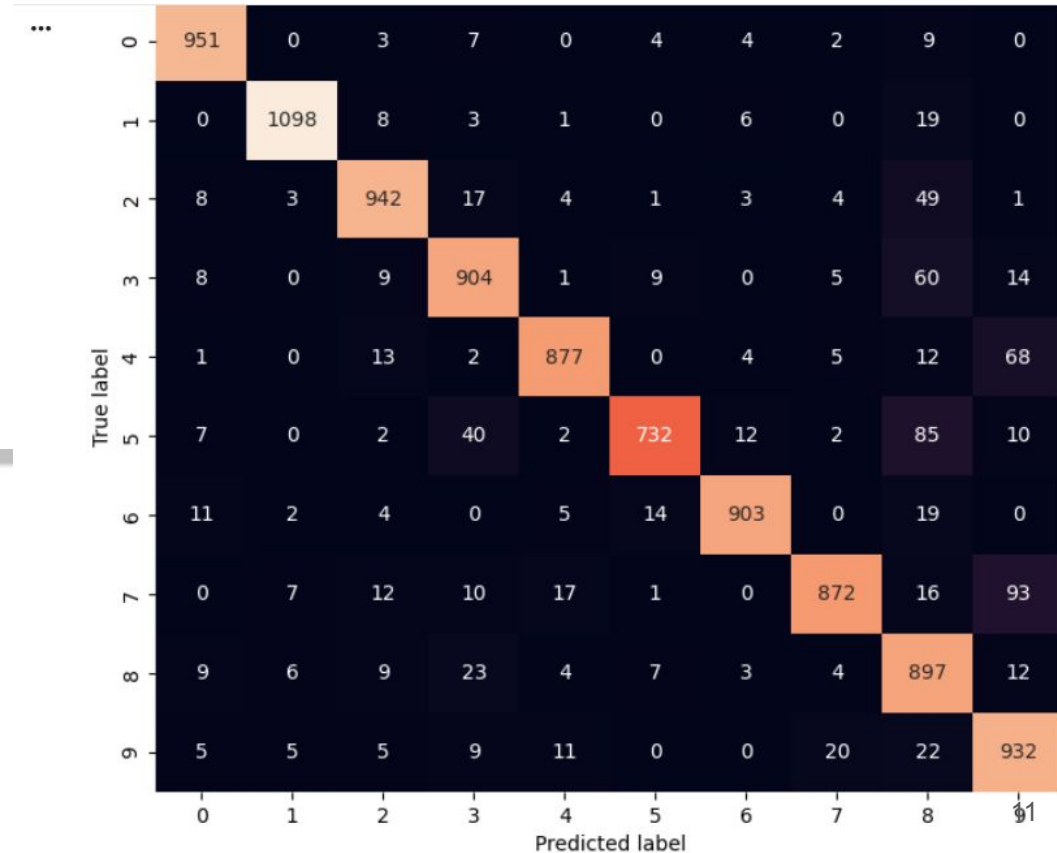
Step 6: Confusion Matrix (Non-Naive Bayes)

This matrix displays the frequency count of each digit in a heatmap format

Min:5, Max:1,

Median:903/904->903.5

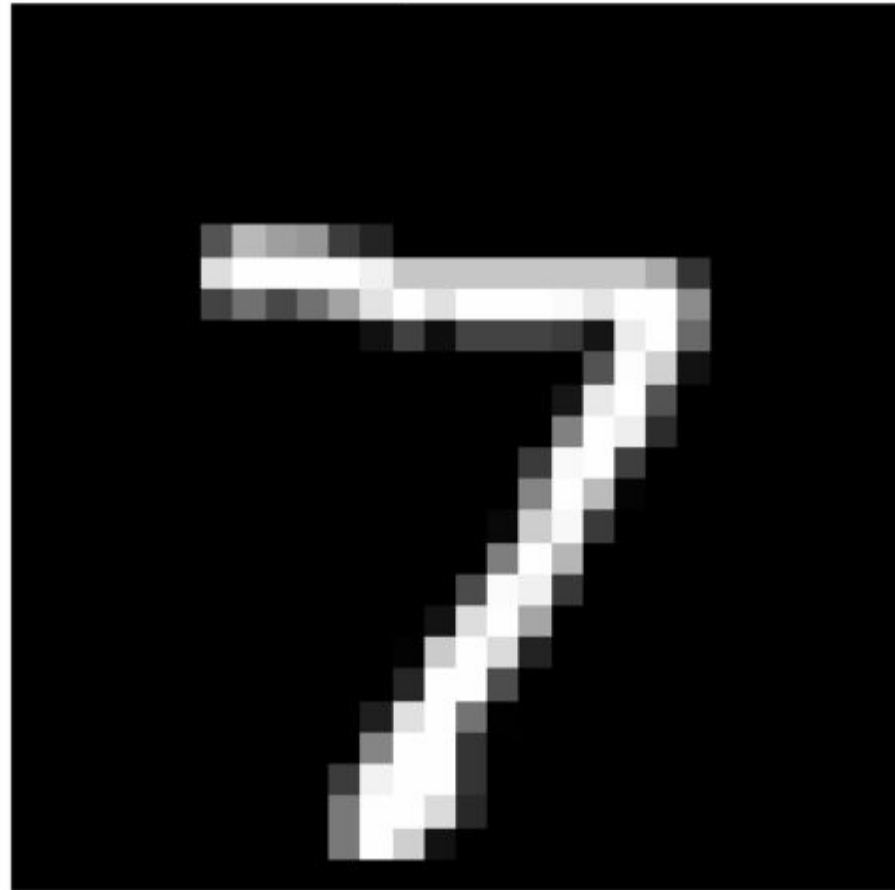
... Model Accuracy: 91.08%



True: 7 | Predicted: 7

Results

- Reshape each image as 28x28 pixel
- Return results of first 10 images



Conclusion

- Automate receipt and tip digitization
- Reduce manual data entry
- Increase accuracy
- Enhance operational efficiency in the food service industry