

Abstract geometric lines in the top-left corner of the slide, consisting of several overlapping, irregular polygons and lines in a light gray color.

# AUTOMATED HANDWRITTEN DIGIT RECOGNITION FOR ENHANCED BETTING SLIP PROCESSING

-BY RAKSHITHA

# INTRODUCTION

2

- 1) As of 2024, the UK gambling industry comprises over 2,500 licensed operators.
- 2) For example, X alone operates approximately 1350 betting shops across the UK. 2700 employees work under company X (Approx 2 in a shop)
- 3) 90% of Bets are placed and translated manually.
- 4) To manage this task, X company employs approximately 50% staff members dedicated to translating and placing these bets efficiently.

# PROBLEM STATEMENT

3

The primary objective is to significantly minimize human errors and revolutionize the manual processing workflow by implementing machine learning algorithms, thereby transitioning towards a fully digitalized and automated system.

# OVERVIEW OF MNIST DATA

- 1) Dataset contains a training set of 60,000 images along with labels.
- 2) A test set of 10,000 images of handwritten digits.
- 3) The handwritten digit images have been size-normalized and centered in a fixed size of  $28 \times 28$  pixels.

# NORMALIZING AND TUNING

5

- 1) The Data set is normalized using scaling them to a range of  $[0, 1]$ .
- 2) Tuning has been handled by altering value of epsilon as  $1e-1$  for getting highest accuracy.

# SELECTING EFFICIENT MODEL

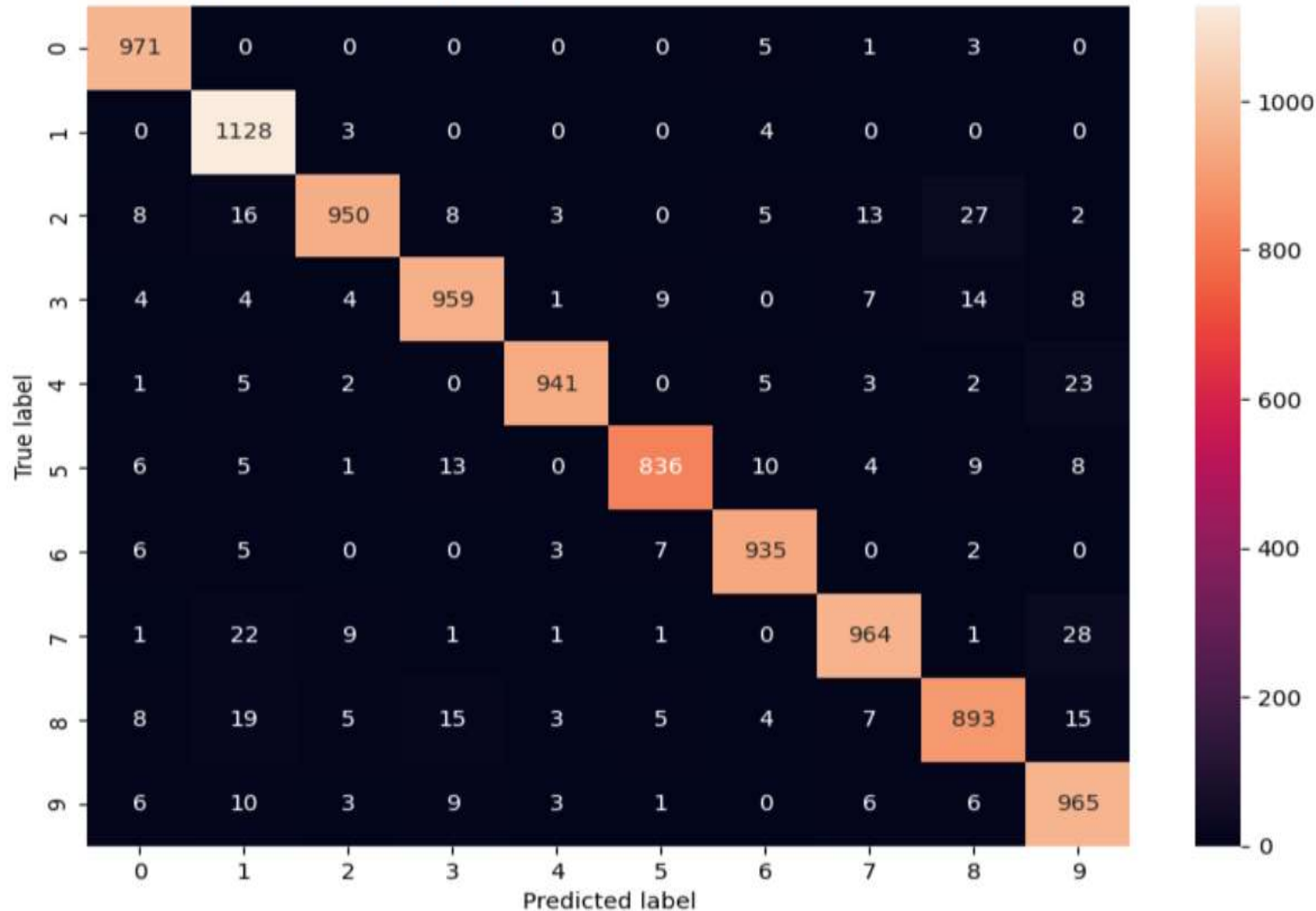
10

	Naïve Bayes	Gauss Bayes	KNN
Before Normalizing and tuning	Train accuracy=59.99% Test accuracy=61.5%	Train accuracy=78.56% Test accuracy=88.5%	Train accuracy=89.8% Test accuracy=89%
After Normalizing And tuning	Train accuracy= <b>77.91%</b> Test accuracy= <b>76.51%</b>	Train accuracy= <b>95.49%</b> Test accuracy= <b>95.42%</b>	Test accuracy= <b>94.94%</b>

# GAUSS BAYES CONFUSION MATRIX FOR ACTUAL LABEL VS PREDICTED LABEL

7

Text(0.5, 47.72222222222222, 'Predicted label')

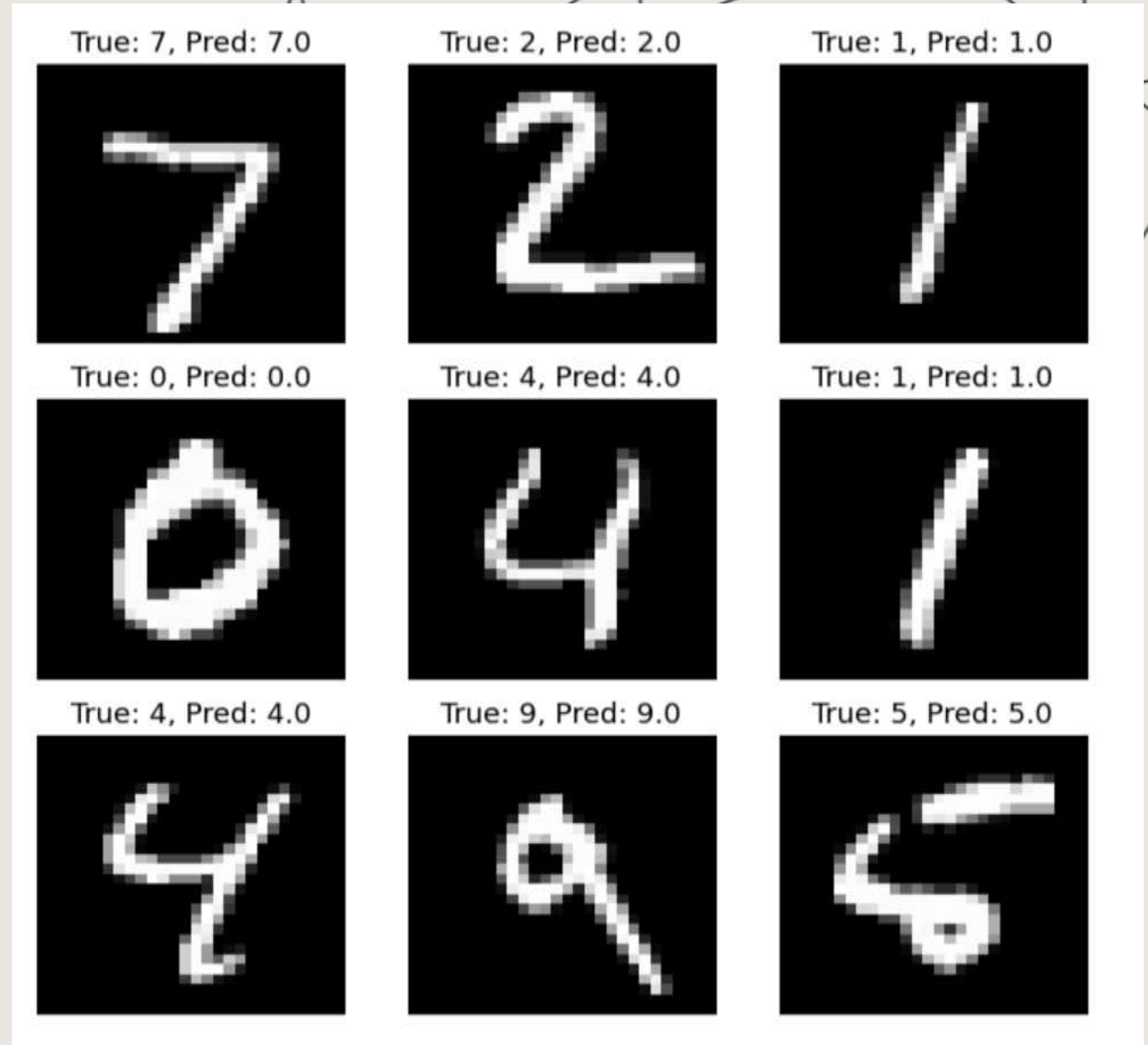


- **True Positives (TP):** The diagonal cells (e.g., 971, 1128) represent the number of correct predictions for each class.
- **False Positives (FP):** The off-diagonal cells in each row (e.g., 5, 3) represent the number of incorrect predictions for each class.
- **False Negatives (FN):** The off-diagonal cells in each column (e.g., 4, 3) represent the number of instances where the model failed to predict the correct class.

# EXAMPLE FOR CORRECTLY PREDICTED VALUES

8

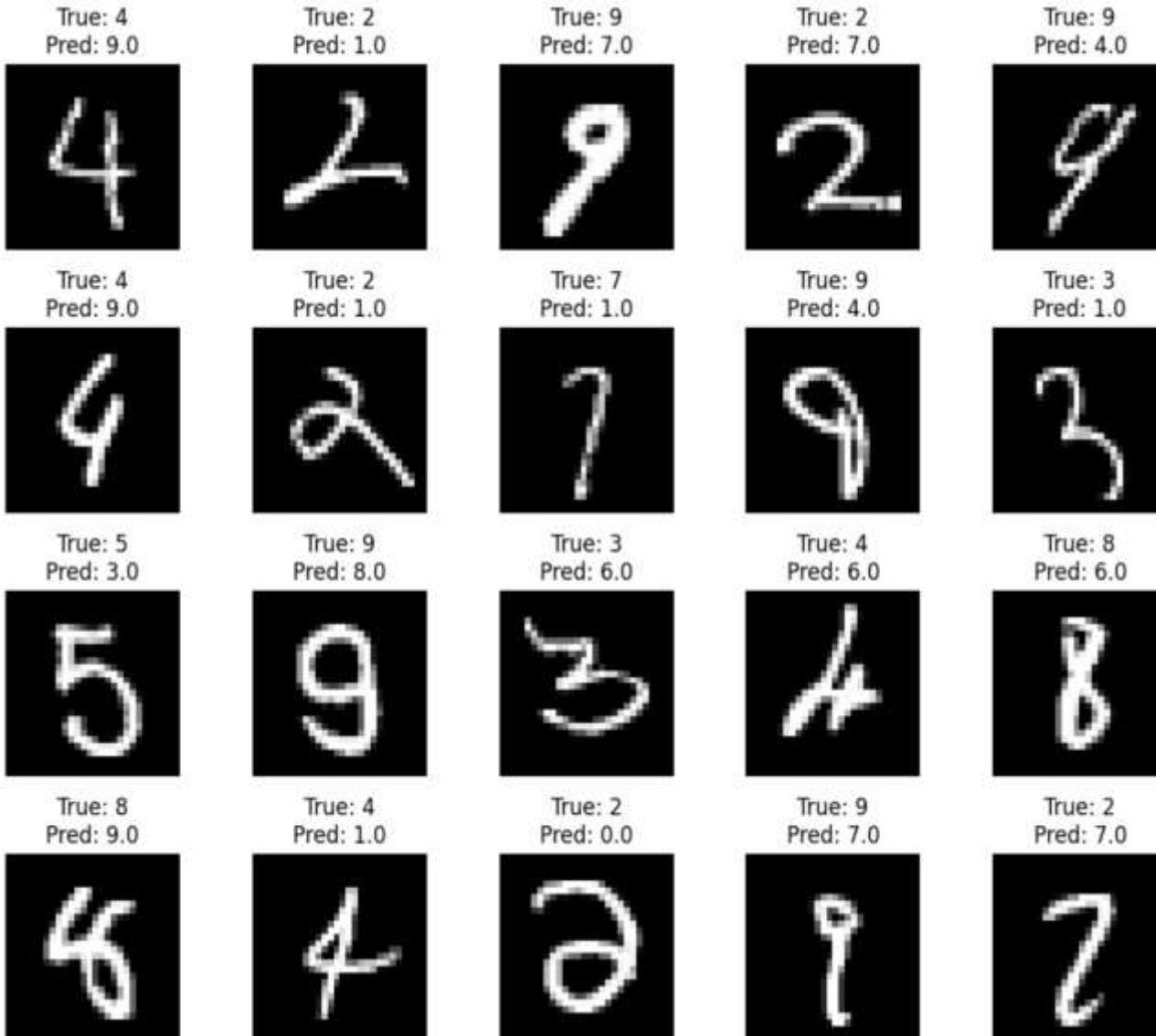
- **High Accuracy:** The model correctly predicted all the digits in the 3x3 grid. Each “True” label matches the “Predicted” label, indicating that the model performs well on this sample set.
- **Consistent Performance:** The model’s predictions are consistent across different digits, from 0 to 9. This suggests that the model has been trained effectively to recognize a variety of handwritten digits.
- **Clear Visualization:** The grid format with true and predicted labels provides a clear and straightforward.





# EXAMPLE FOR INCORRECTLY PREDICTED VALUES

9



- The machine learning model incorrectly predicted several digits, such as predicting a “9” instead of a “4” and a “1” instead of a “2”.
- The model’s errors are consistent across different digits, indicating less accuracy and not 100%.

## Typical Betting slips



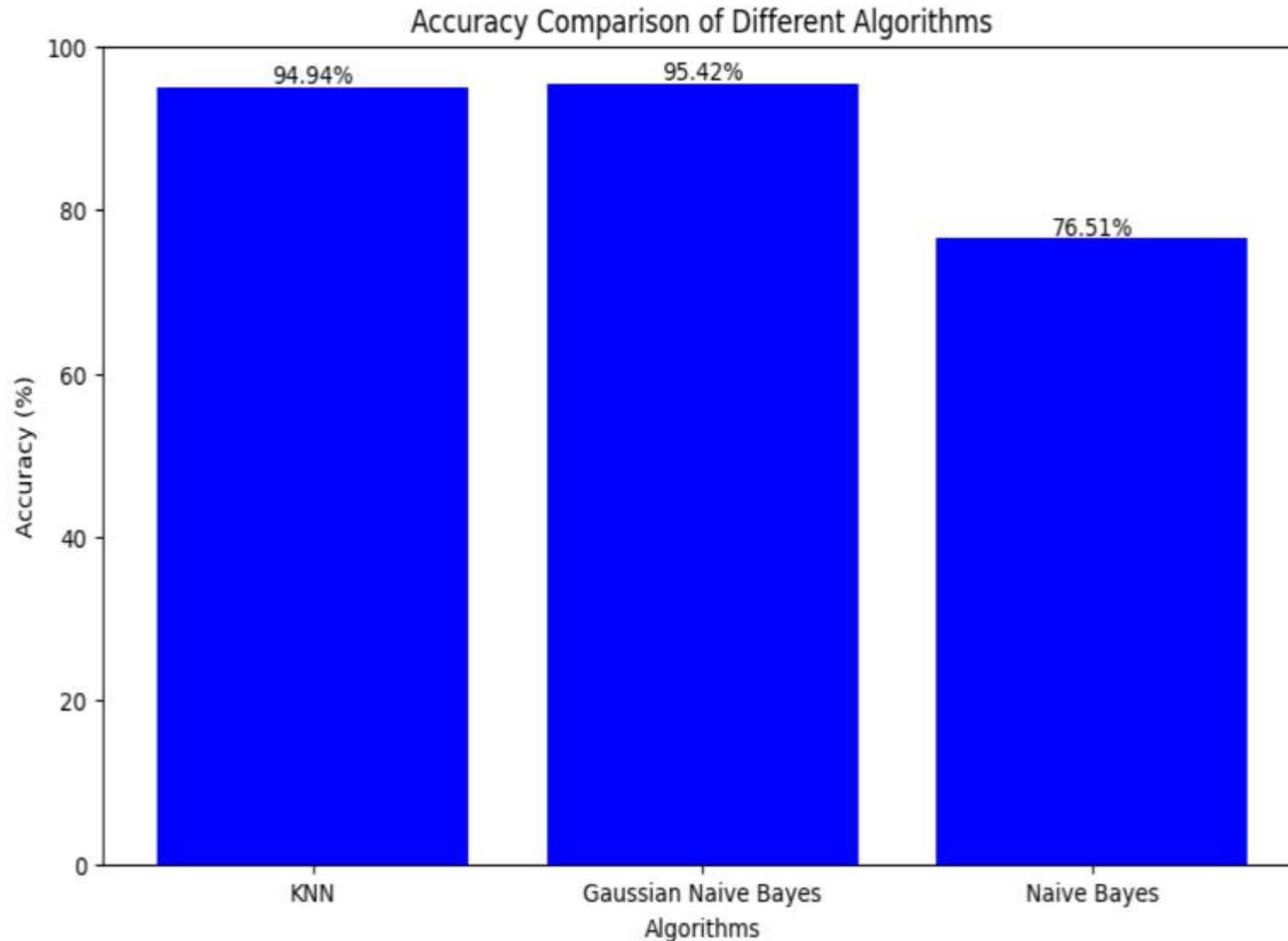
## Processed slip using KNN Preprocessed Image



Predicted Label  
using KNN=1

# COMPARISON BETWEEN DIFFERENT MODELS

11



- The bar graph compares the accuracy of three algorithms: KNN, Gaussian Naive Bayes, and Naive Bayes.
- KNN and Gaussian Naive Bayes have similar high accuracy rates, around 94.94% and 95.24% respectively.
- Naive Bayes has a significantly lower accuracy rate of approximately 76.51%, indicating it performs worse than the other two algorithms in this comparison.

# SUMMARY

- 1) Gauss Naïve bayes has highest accuracy.
- 2) By implementing this system, we can enhance betting system and reduce cost to the company.
- 3) Further, helps to move eco-friendly by reducing paper usage.
- 4) Digital transformation of betting system.

# APPENDIX

1) Why epsilon value  $1e-1$  gives maximum accuracy ?

For the MNIST dataset, the epsilon value of  $1e-1$  likely provides an optimal balance between regularizing the covariance matrix and maintaining model performance. This helps to stabilize the matrix while allowing the model to fit the data well, thus achieving maximum accuracy.

2) What library is used to convert image while processing?

Pillow is the library which was used to process image.

A series of white, overlapping geometric lines and polygons on a black background, located on the left side of the slide.

# THANK YOU

Rakshitha Murali Krishna

[Rakshitha.krishna@informationtechconsultants.co.uk](mailto:Rakshitha.krishna@informationtechconsultants.co.uk)