HANDWRITTEN DIGIT RECOGNITION

CASE STUDY FOR HOSPITALS
(DOCTORS DIGIT WRITING PREDICTION)

28 October, 2024

HELLO!

Warm greetings to all present. Been here today, I am excited to introduce my case study for ***Hospitals (Doctors Handwritten Digit Prediction),*** which aims to address key challenges and capitalize on emerging opportunities in the market.

AGENDA

01

BACKGROUND OF THE STUDY

02

PROBLEM STATEMENT

03

DATASET OVERVIEW

04

ANALYSIS

05

FEATURE ENGINEERING

06

MODEL BUILDING

07

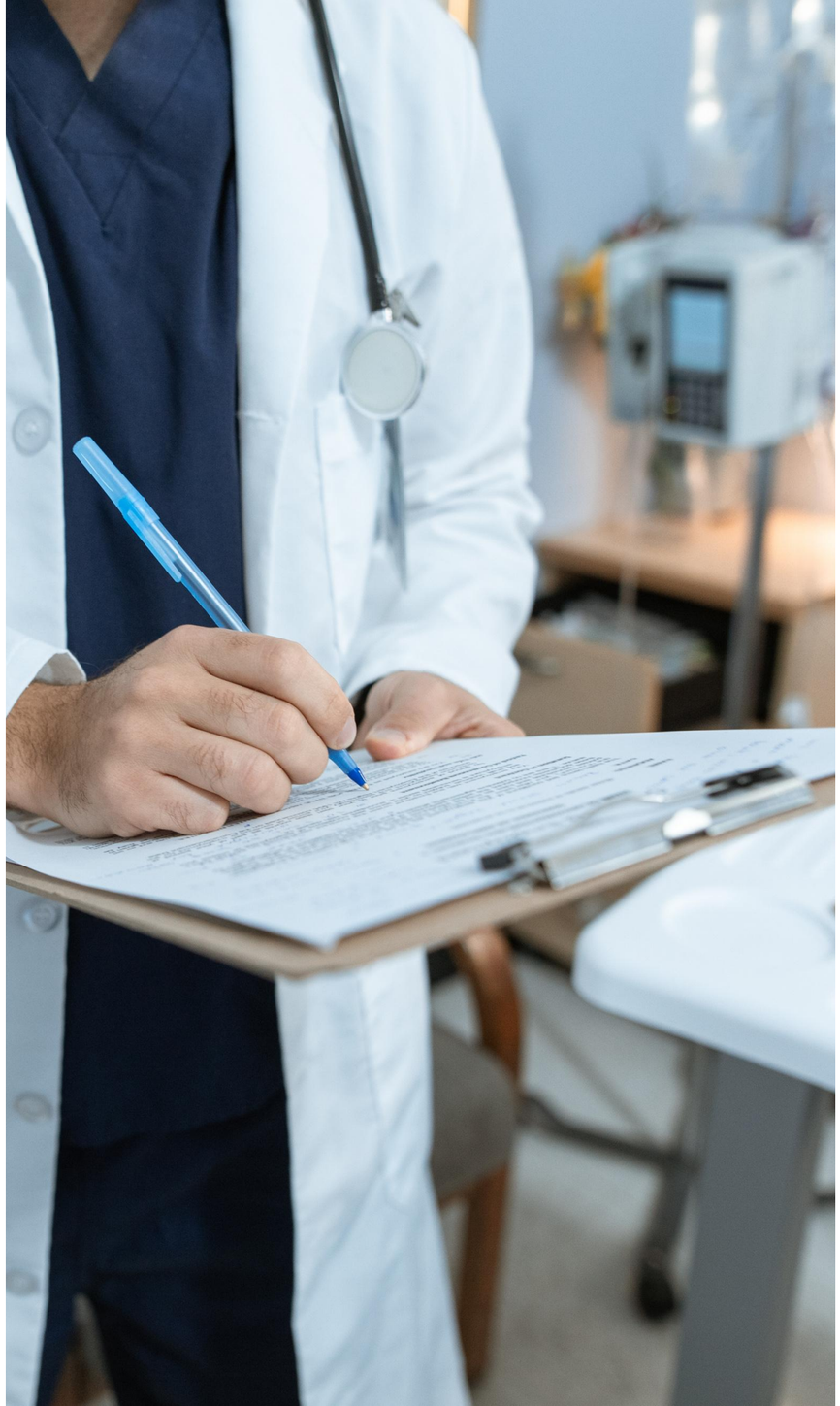
RESULTS

08

CONCLUSION

Doctors don’t have a bad handwriting they just have their own unique style.

-- Vinay



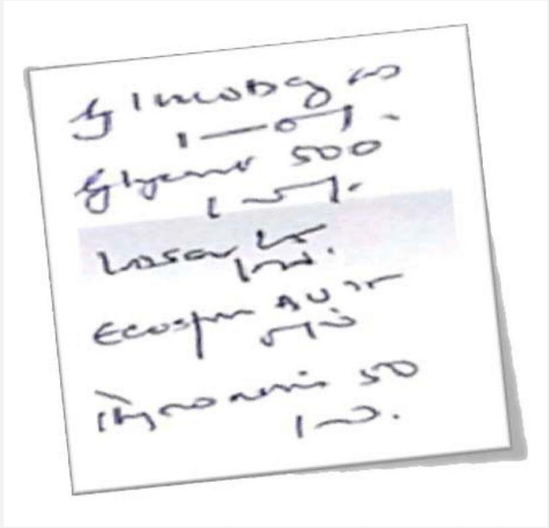
BACKGROUND OF THE STUDY

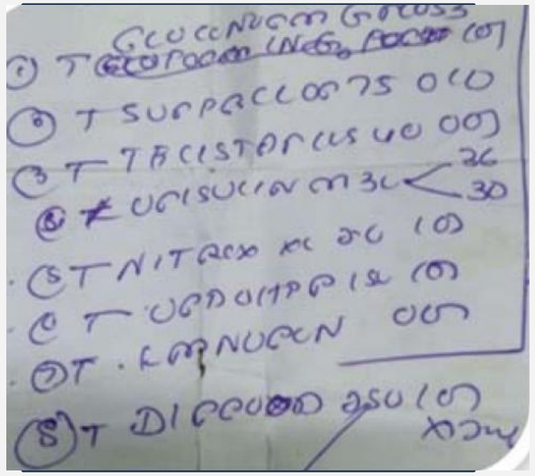
- Handwritten notes from doctors are critical for patient records and communication.
- Misinterpretations can lead to errors in critical medical information
- In USA **44,000 – 98,000** people die annually from medical errors. Out of them, about **7,000 deaths** are attributed to sloppy handwriting.
- In UK, about **30,000 British die** every year due to medical errors.

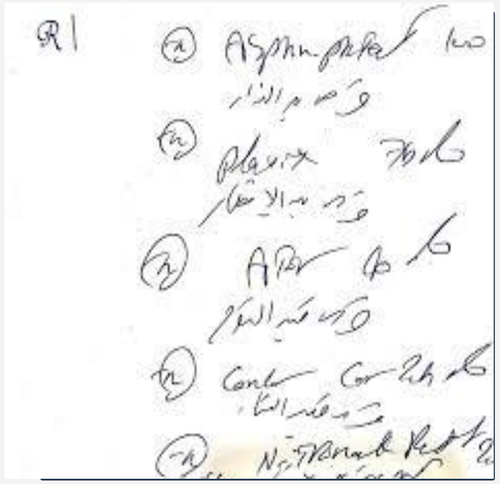
--- In 1999, the Institute of Medicine (IOM), USA, released a study report To err is human: Building a safer Health System.

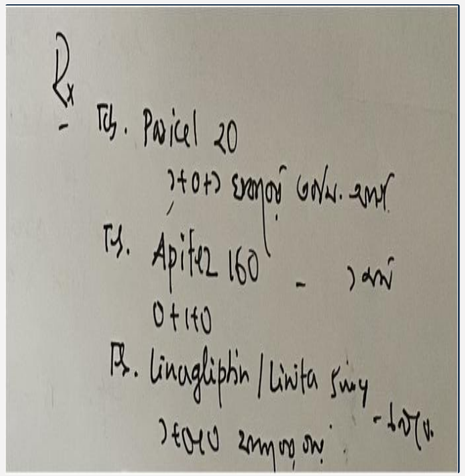
https://www.deccanchronicle.com/nation/in-other-news/201018/its-time-to-totally-ban-handwritten-prescription.html

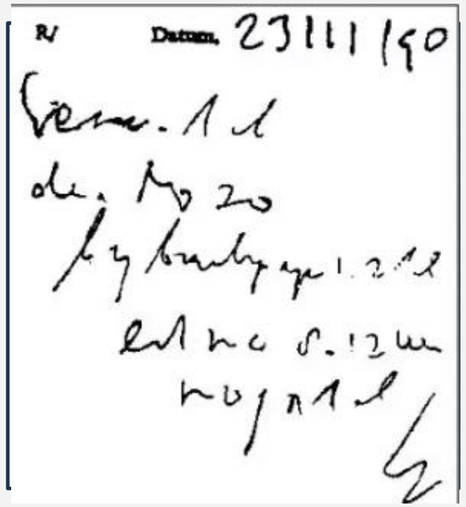
Images

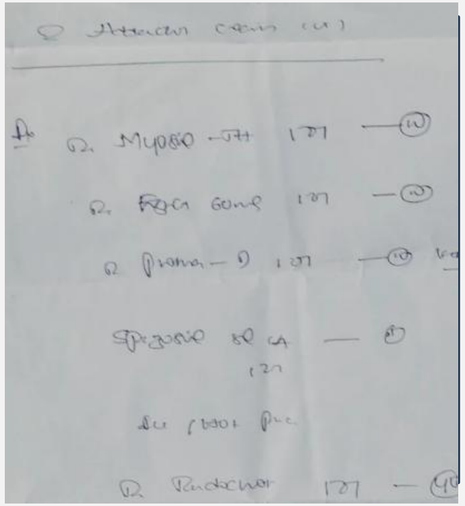












References has given in the notes

PROBLEM STATEMENT

Our use case focuses on accurately recognizing handwritten digits in doctors' notes to improve patient record accuracy and reduce errors in hospitals.

01

SCOPE OF THE STUDY

Our study aims to pinpoint challenges and opportunities, assess feasibility, and offer strategic insights.

02

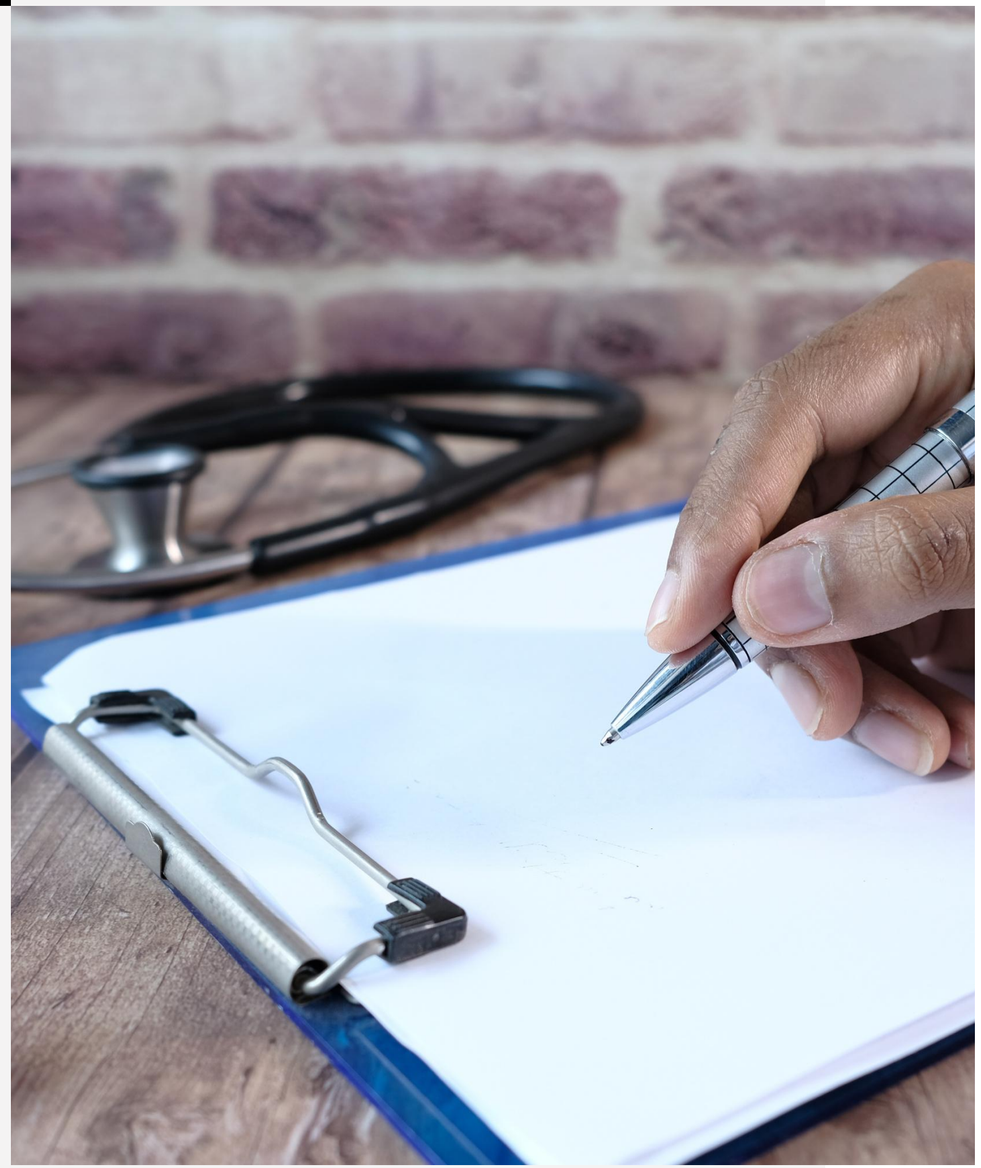
RELEVANCE OF THE STUDY

Improving data accuracy in hospitals, reducing errors, and enhancing patient safety in healthcare documentation.

03

RESEARCH QUESTION

How can machine learning improve the accuracy of interpreting handwritten digits in doctors' notes to enhance patient data reliability and safety in hospitals



DATASET OVERVIEW

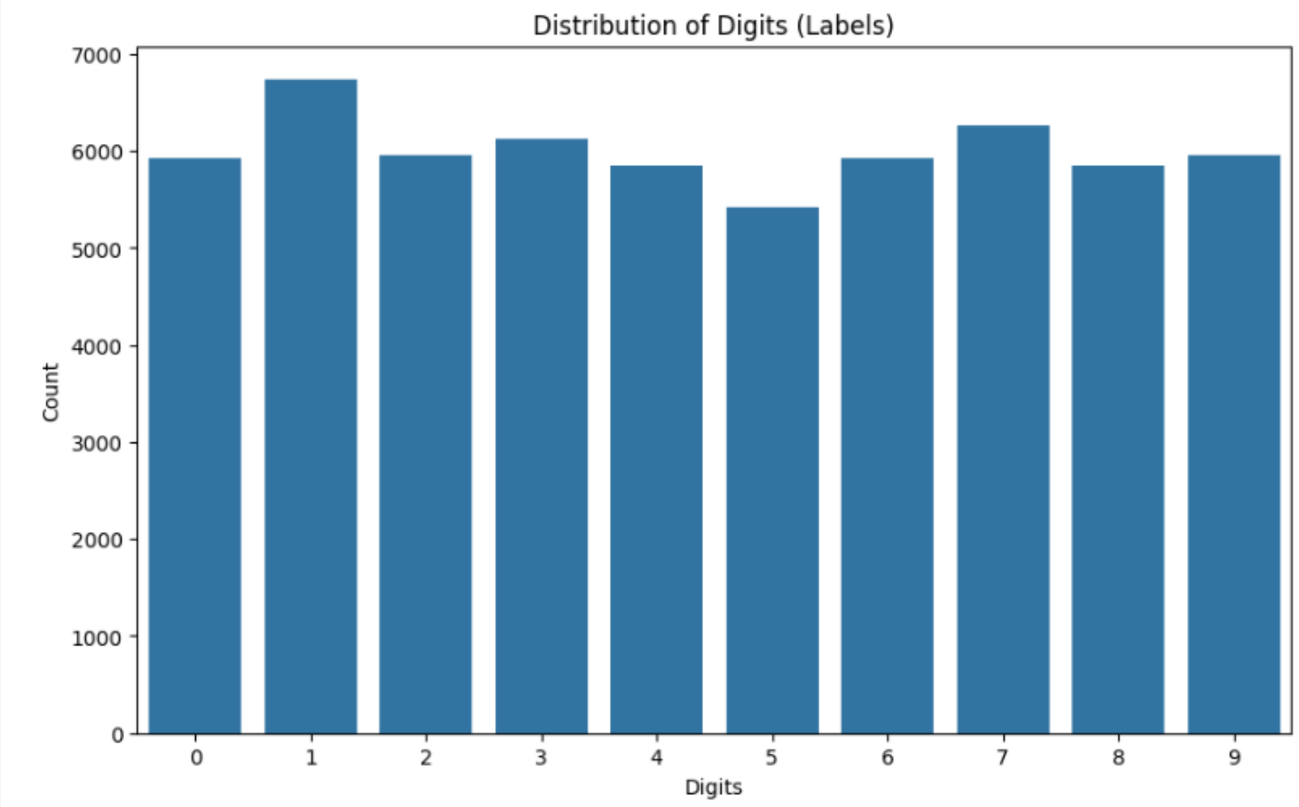
This dataset consists of mnist pixels which comprises of few features which includes; Labels and pixels (numbers - 785 data numeric points).

DATASET SIZES:

- Training Data Shape: (60000, 787)
- Test Data Shape: (10000, 787)

- Each row consists of 785 values: the first value is the label (a number from 0 to 9) and the remaining 784 values are the pixel values (a number from 0 to 255).

STATISTICAL & DESCRIPTIVE ANALYSIS



This effectively highlights the count of expected predicted numbers in the dataset given which is ok for training a perfect model.

MISSING VALUES, NORMALIZATION AND STANDARDIZATION OF DATASET

01

MISSING VALUES

Our dataset has no missing values but we had to drop some un-needed columns for model performance.

02

NORMALIZATION

We implemented data normalization by implementing:

X_train_norm = X_train / 255

X_test_norm = X_test / 255

to scales the values to a range between 0 to 1.

03

STANDARDIZATION

The data was standardized to optimize data performance for predicting with best accuracy.

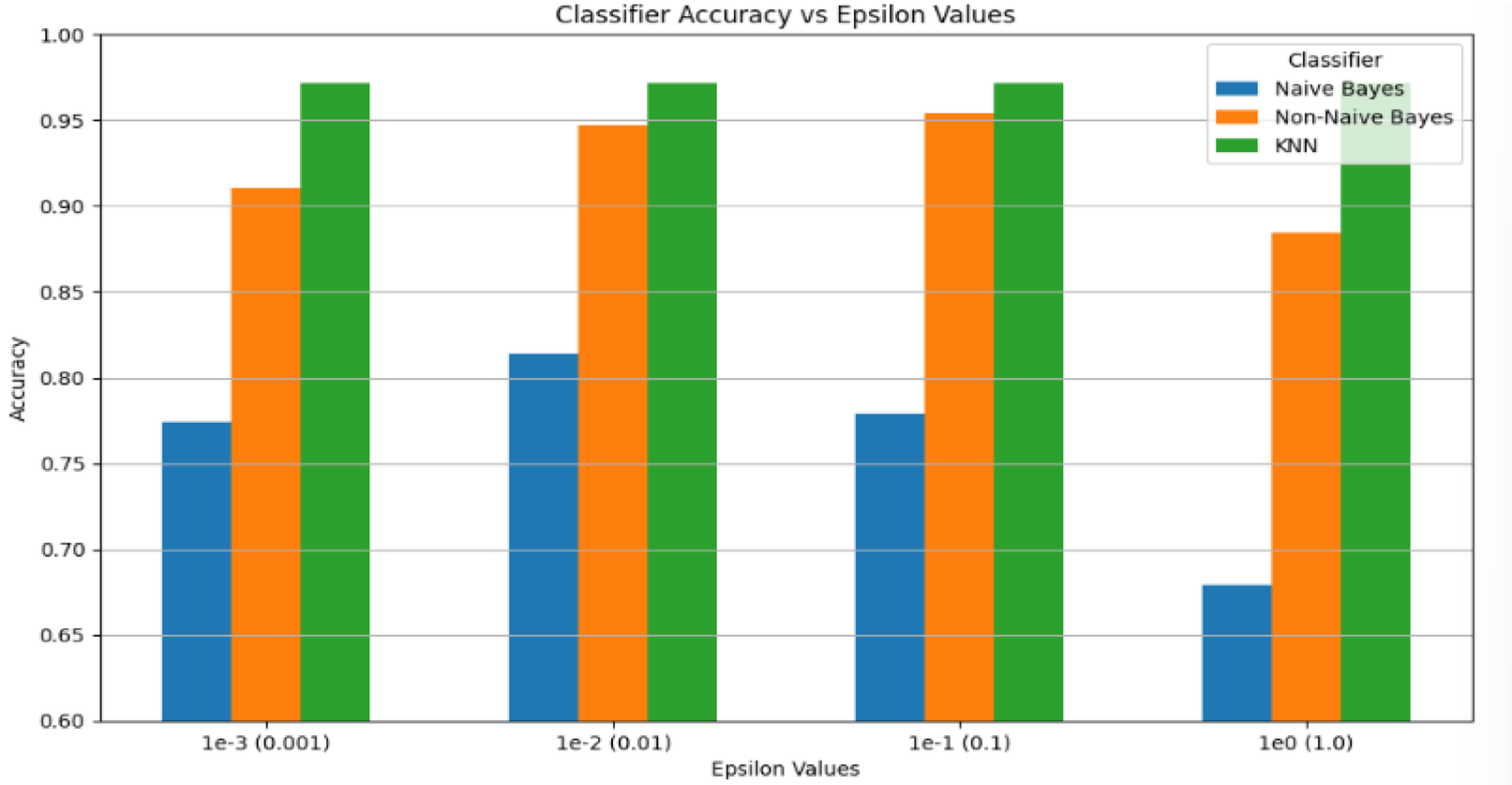
MODEL BUILDING

CLASSIFIER IMPLEMENTATIONS:

1. **NAIVE BAYES:** It calculates the probability of each class given the features and selects the class with the highest probability.
2. **NON- NAÏVE BAYES:** It doesn’t assume independence among features when making predictions. They leverage complex relationships and interactions between features to improve classification accuracy.
3. **K-NEAREST NEIGHBORS WITH DISTANCE WEIGHTING:** A straightforward classifier that looks at the closest points to decide the class of a new point.

TESTED WITH EPSILON VALUES OF [1E-3, 1E-2, 1E-1, 1E0]

CLASSIFIER ACCURACY & EPSILON VALUES



EPSILON	NAÏVE BAYES	NON-NAÏVE BAYES	KNN
1e-3 (0.001)	0.7746	0.9108	0.9717
1e-2 (0.01)	**0.8141**	0.9473	0.9717
1e-1 (0.1)	0.7791	**0.9542**	0.9717
1e0 (1.0)	0.6791	0.8847	0.9717

RESULTS

- Epsilon Impact on Accuracy
  - Naïve Bayes, accuracy is highest at 81% but then decreases at higher epsilon values at 0.01
  - Non-Naïve Bayes is highest at 95% maintaining a higher performances across epsilon values at 0.1
  - KNN consistently accurate at 0.97 %

- Classifier Comparison
  - KNN performs both naïve bayes and non naïve bayes overall accuracy across all epsilon values

CONCLUSION

- This study successfully implemented machine learning classifiers to enhance the accuracy of handwritten digit recognition in doctors' notes within hospital settings.
- KNN significantly improved handwritten digit recognition accuracy in doctor’s notes
- Reduced errors in digit interpretation.
- Exploring advanced models, such as deep learning, could further increase more accuracy

AFTER IMPLEMENTIG THIS APPROACH FOR HOSPITALS

EXPECTED ATTRIBUTES

THANK YOU

--- I solve your toughest challenges with the power of AI and machine learning.

28 October, 2024