# Devanagari - Handwritten Character Classification

Sheetal Munjewar

Bellevue University

Course Number: DSC 680 Applied Data Science

Amirfarrokh Iranitalab

02/28/2025

https://smunjewar.github.io/Portfolio/

# About Devanagari

**Origin**: Devanagari evolved from the Brahmi script around the 7th century CE and is often referred to as the "script of the gods."

**Meaning**: The name "Devanagari" comes from the Sanskrit words *Deva* (god) and *Nagari* (city), highlighting its association with Sanskrit and religious texts.

**Structure**: The script features 33 consonants and 14 vowels. Vowels are written independently at the beginning of syllables, while consonants have inherent vowels that can be altered with diacritics.

**Horizontal Line**: Devanagari is written from left to right, with a distinctive horizontal line running across the top of each word, connecting letters.

**Languages**: It is used to write several languages, including Hindi, Sanskrit, Marathi, Nepali, and Konkani, among others.

**Literature and Media**: Devanagari is widely used in literature, newspapers, books, films (especially Bollywood), and radio across South Asia.

**Digital Use**: The script is supported by Unicode, making it common in digital platforms, including websites, social media, and emails.

**Religious Significance**: Devanagari is the script for many Hindu, Buddhist, and Jain scriptures and is used for chanting, prayers, and religious rituals.

**Educational Importance**: It is taught in schools across South Asia for languages such as Hindi, Sanskrit, and Marathi, playing a key role in linguistic and cultural education.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| क | ख | ग | घ | ङ | च | छ | ज | झ | ञ |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| ट | ठ | ड | ढ | ण | त | थ | द | ध | न |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| प | फ | ब | भ | म | य | र | ल | व | श |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| ष | स | ह | क्ष | त्र | ज्ञ | | | | |
| 31 | 32 | 33 | 34 | 35 | 36 | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| English Numerals | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Devanagari Numerals | ० | १ | २ | ३ | ४ | ५ | ६ | ७ | ८ | ९ |

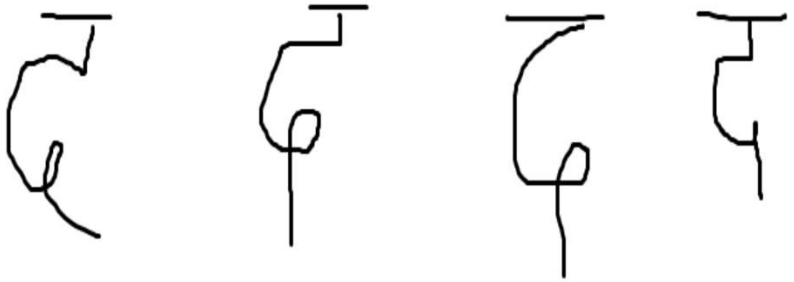
# Problem Statement

- Offline Handwritten Recognition
- Complexity

Similar letters

क फ
ट ठ ढ द
व ब

Different styles of writing



## Steps Performed

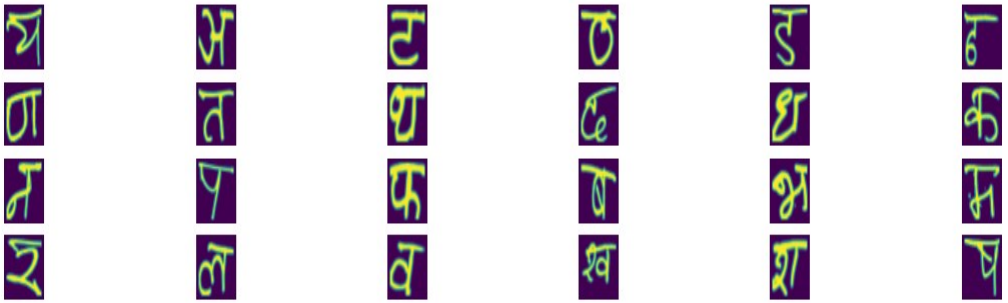
- Data gathering
- Data understanding
- Data preparation
- Model Compilation
- Model Training
- Testing model

# Dataset

- Source: Center for Machine Learning and Intelligent Systems
- 46 classes
- 2000 images per class
- Dataset split in 85-15
- 32x32 image size

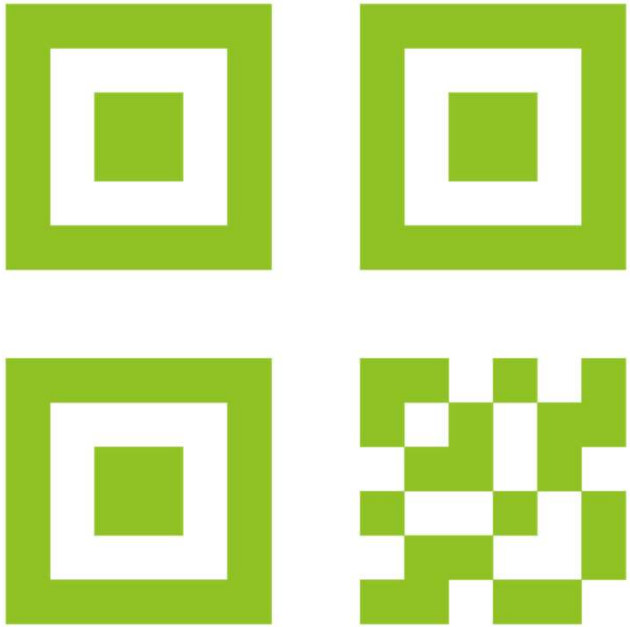
# Data Understanding

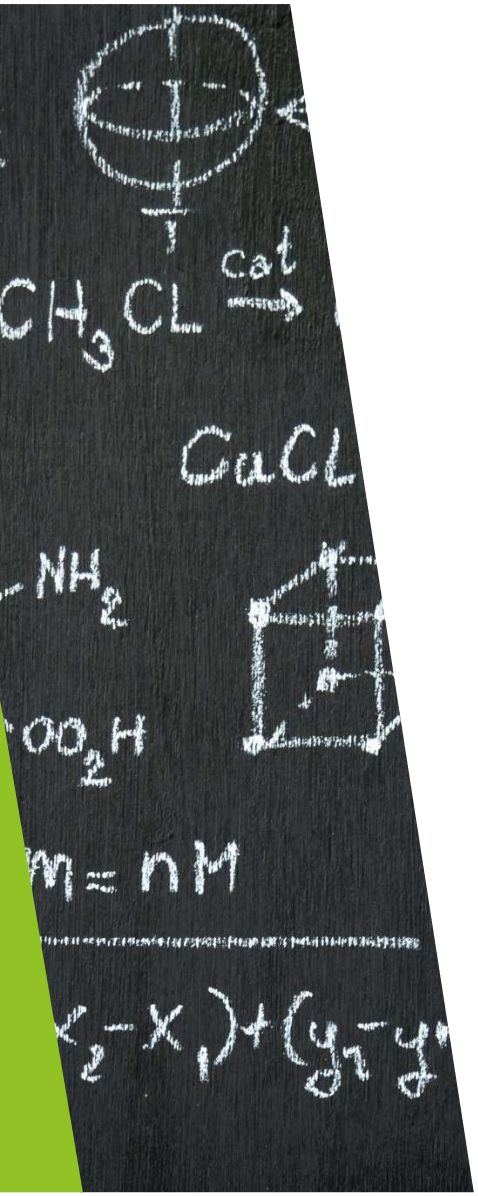
- Folder Structure
- 92,000 images
- 78,200 images in Training dataset
- Equal samples for each character



# First Approach - Machine Learning

- Scikit-learn Models
- Data frame with 1025 columns
- 1024 pixel columns
- Normalization
- Label Encoding



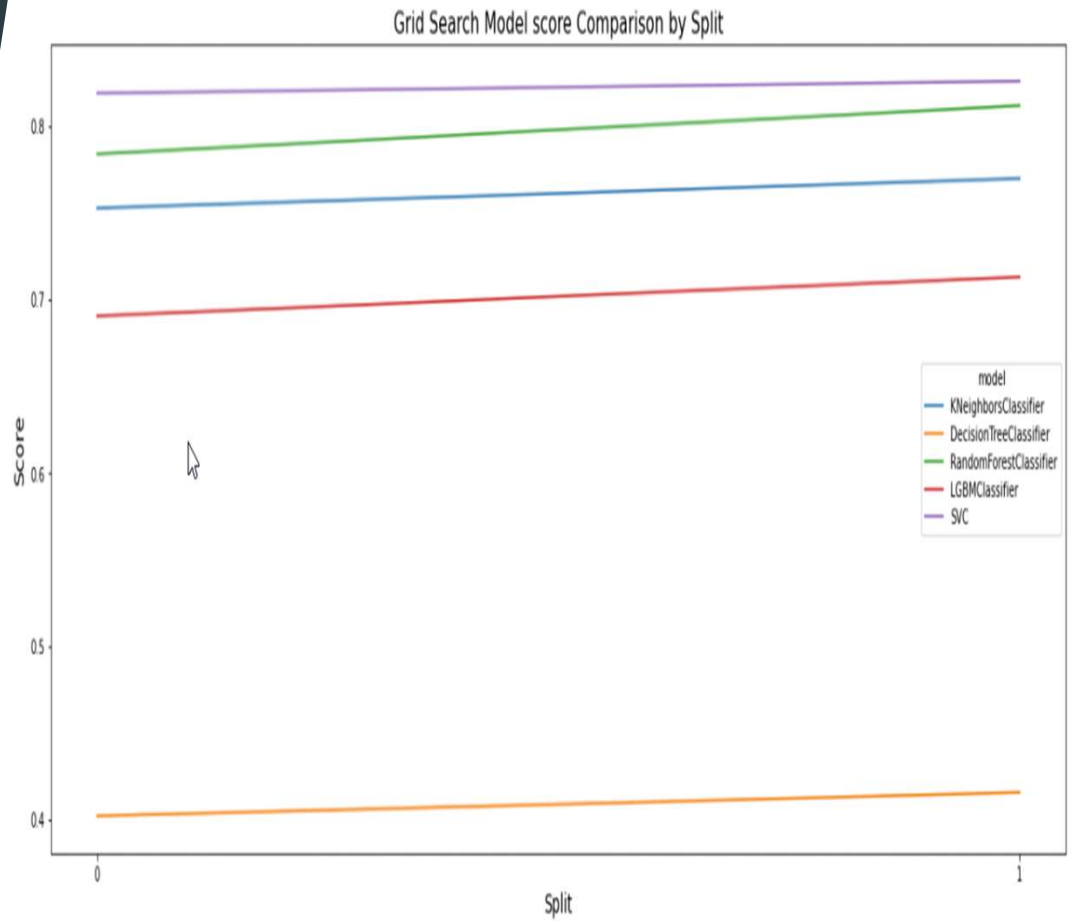


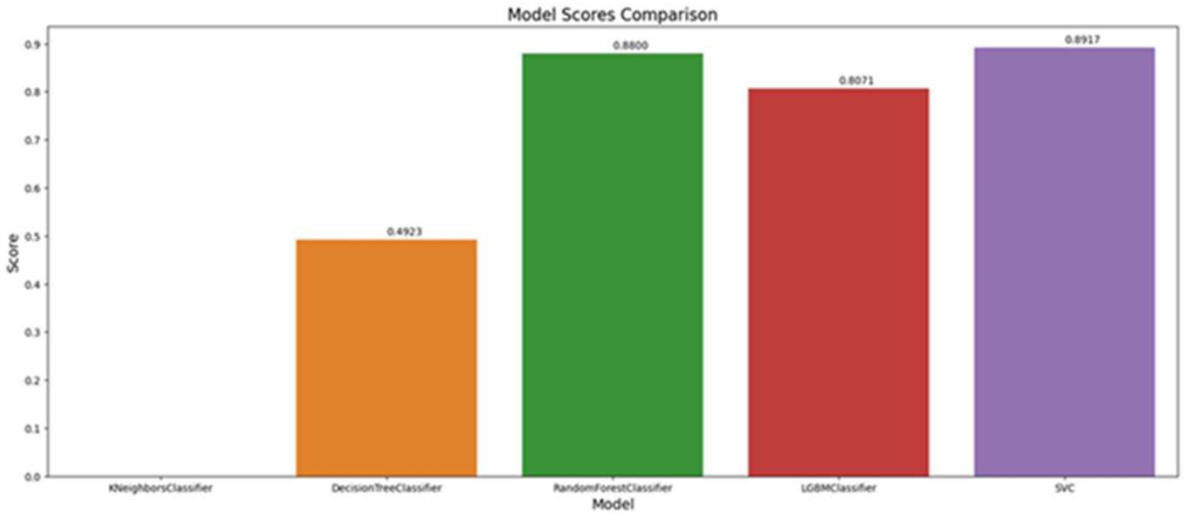

# Machine Learning

- Grid search
- KNeighborsClassifier, DecisionTreeClassifier, RandomForestClassifier, LGBMClassifier, SVM
- 9 + 3 + 3 + 18 + 6 = 39 Models Trained
- Problem with: GradientBoostingClassifier, XGBClassifier

# Machine Learning

- Grid search



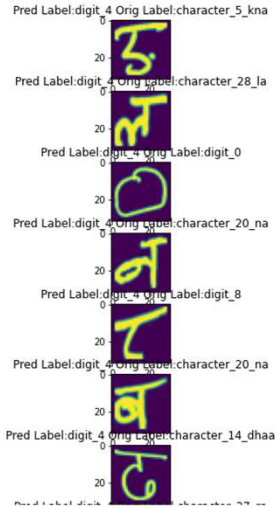


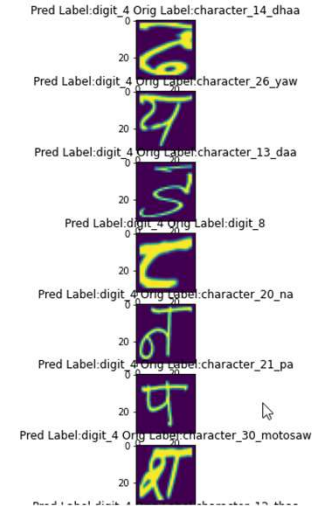

# Machine Learning

Grid search

# Machine Learning

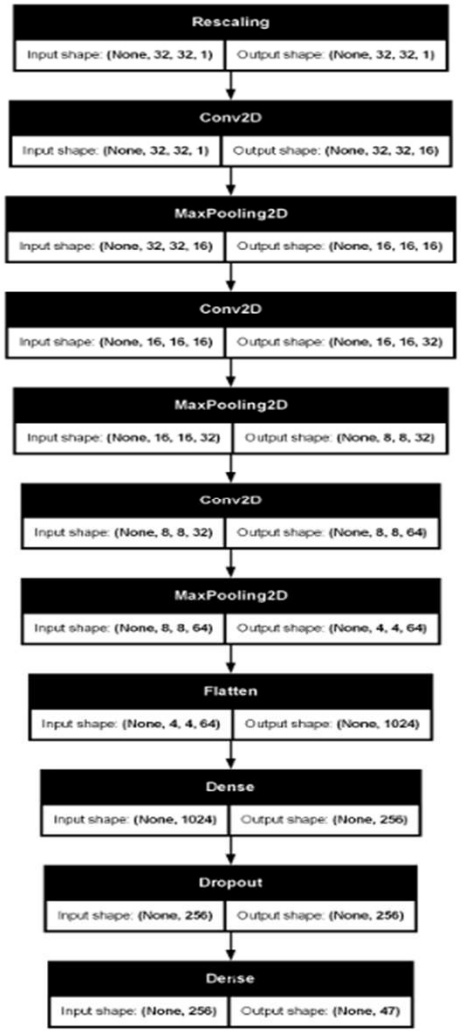
- Best Model: SVM
- Best Model Accuracy on Test Dataset: 64.70%



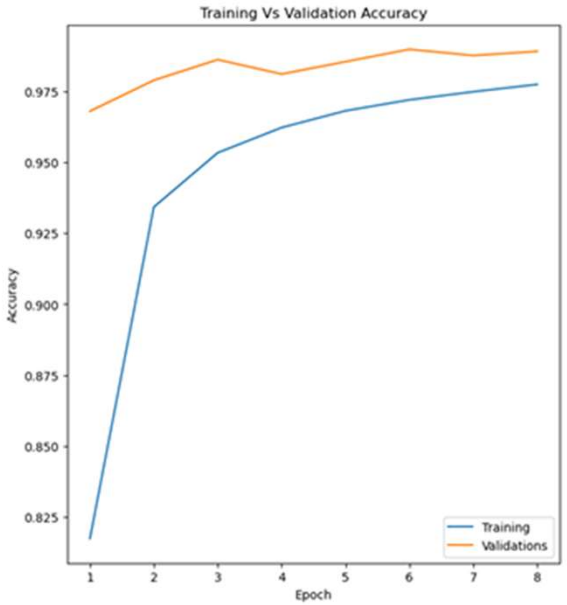
# Seacond Approach - Deep Learning

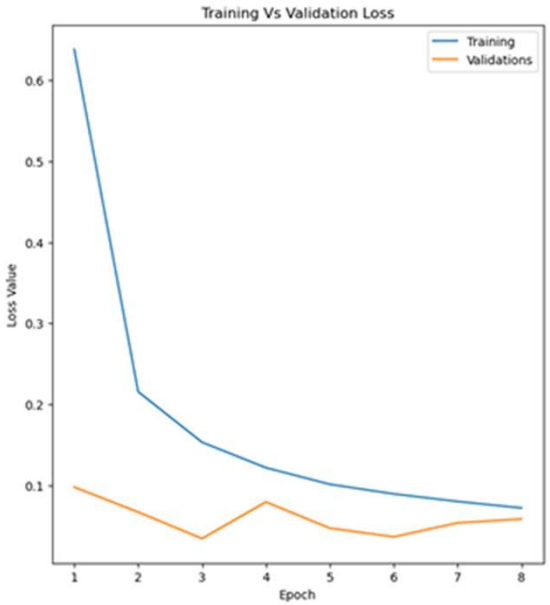
- Tensorflow CNN
- Augmentation
- Vocabulary of labels
- TensorDataset
- Dataset Split
- Batches of 32




# Deep Learning

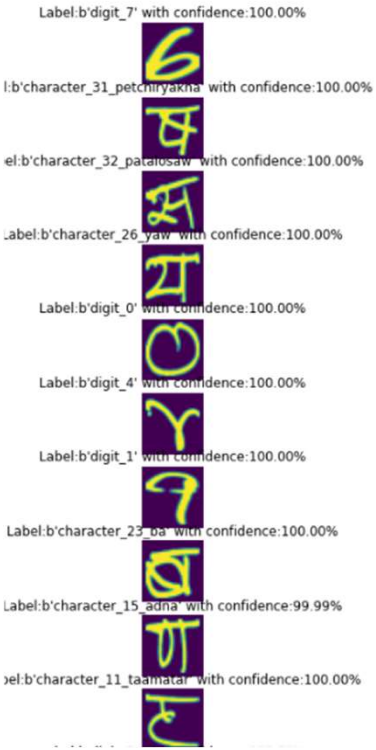
- Define Model
- Compile Model

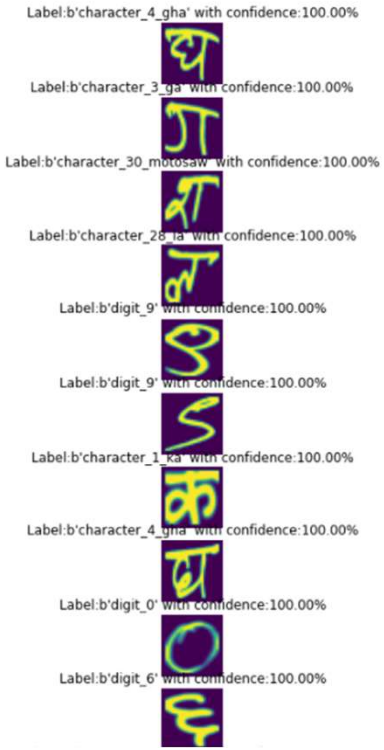



# Deep Learning

Model Training




# Deep Learning

Model Score on Test: 98.91%
with 0.05 Loss

# Thank You!