SUPERVISED LEARNING WITH EMNIST

Cameron Wood

AGENDA

Problem Statement

ML Approaches

Results

Summary

GitHub Repo

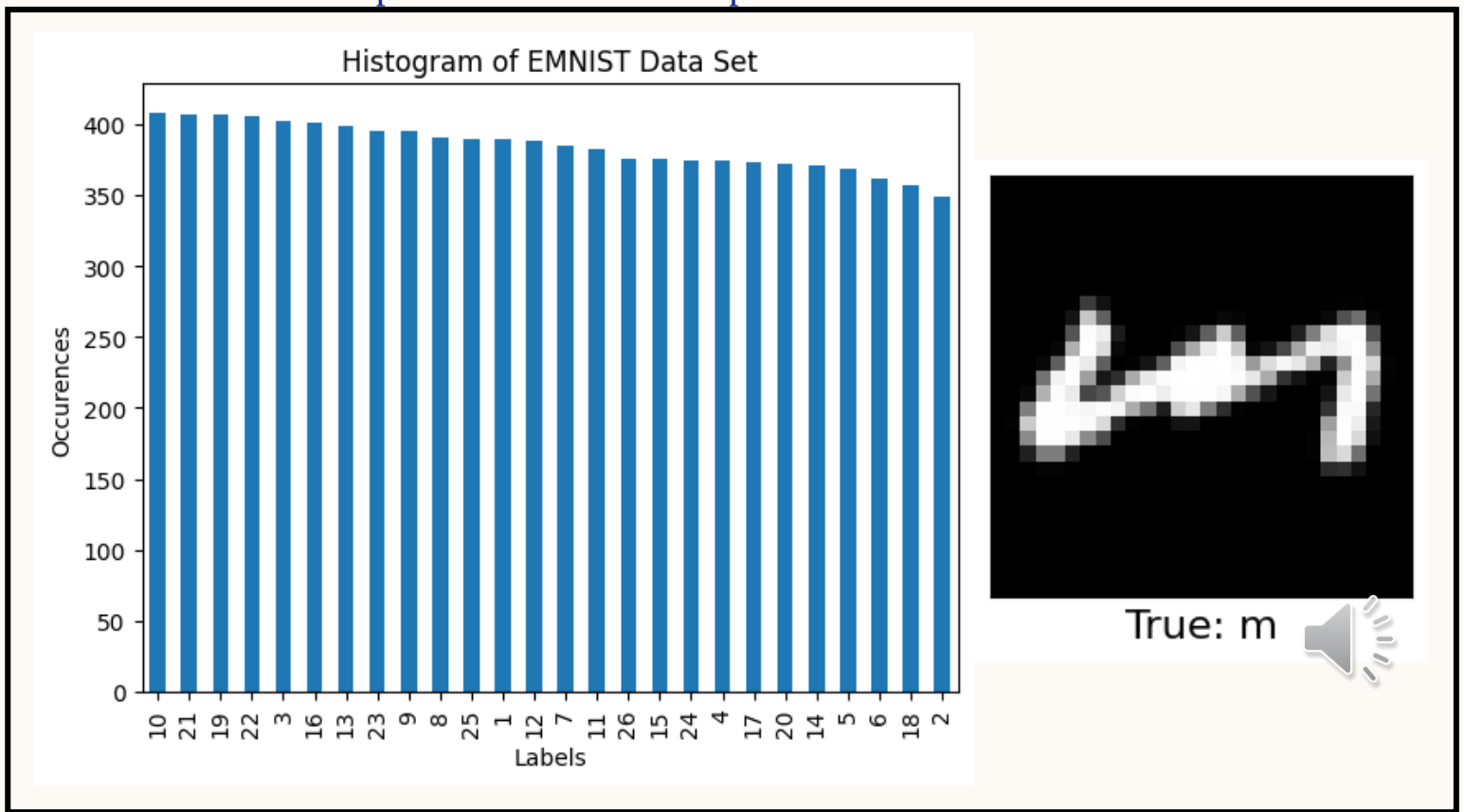
PROBLEM STATEMENT

Image Classification is becoming increasingly important in many industries, including: Medicine, Agriculture, Autonomous Vehicles, Natural Language Processing, and many more. One specific use case this project will be focusing on is the classification of handwritten letters.

Often times this type of classification uses computationally intensive Machine Learning models, like Convolutional Neural Nets. This project aims to show that it's possible to use basic, supervised learning techniques to train high performing models, which can be ran on everyday computers.

EMNIST: AN OVERVIEW

This project will be using the EMNIST Letters Dataset, which contains over 88,000 28x28 handwritten letters. This data was compiled from the NIST Special Database 19.



MACHINE LEARNING MODELS

Logistic Regression and K-Nearest Neighbor

MACHINE LEARNING MODELS USED

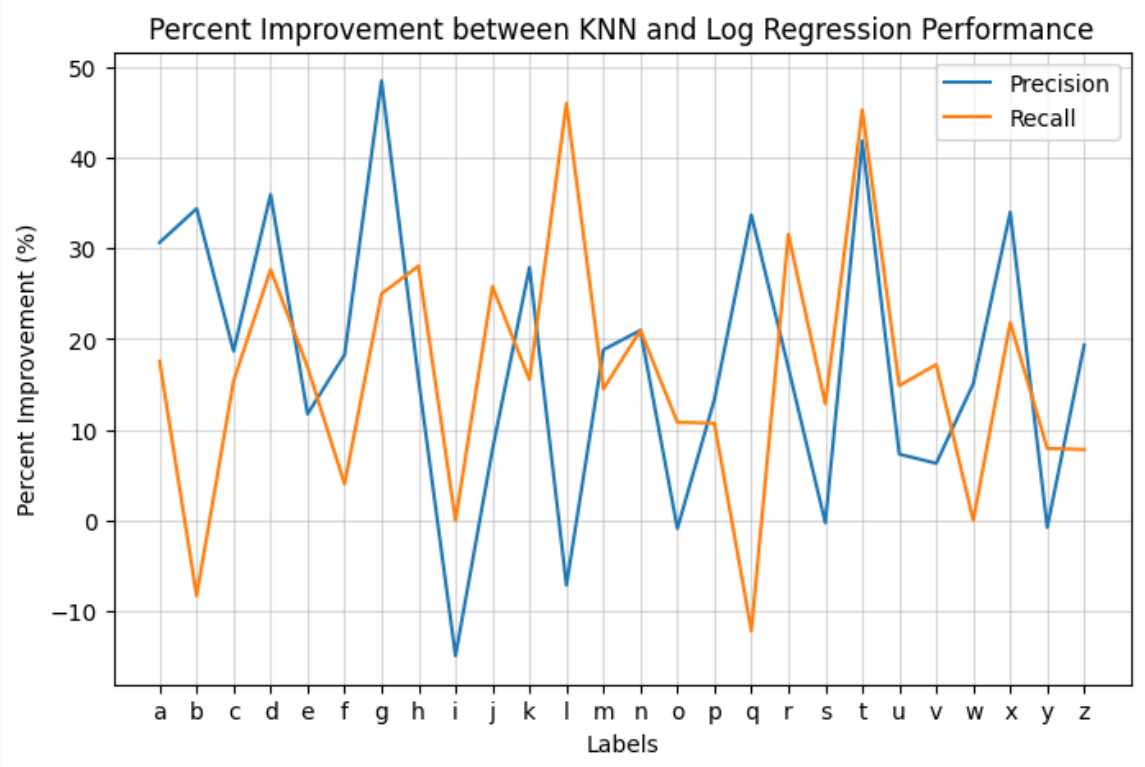
In this project, I chose to use Logistic Regression and K-Nearest Neighbors as these are relatively simple Classification Models perfect for this dataset. They allow for quick iterative designs, and each has their own unique outputs that give great insights into misclassifications to further re-fine the model.

As we'll see in the results, KNN does outperform Logistic Regression due to it's greater flexibility. KNN is also better in situations where features are clustered, which is true for EMNIST where the features correspond to pixel intensity values.

RESULTS

Comparing Models and Overall Performance

PERFORMANCE SUMMARY



Model	Accuracy	Average Recall	Average Precision
Logistic Regression	63.3%	63.5%	63.2%
KNN	75.6%	75.9%	77.7%

SUMMARY

TAKEAWAYS

- Able to achieve up to 76% accuracy with minimal tuning
- Models are easily developed and iterated upon
- Allows for models to be developed on everyday computers and used on a larger scale

THANK YOU!