

# Русский-CV

Optical Character Recognition with Russian Characters

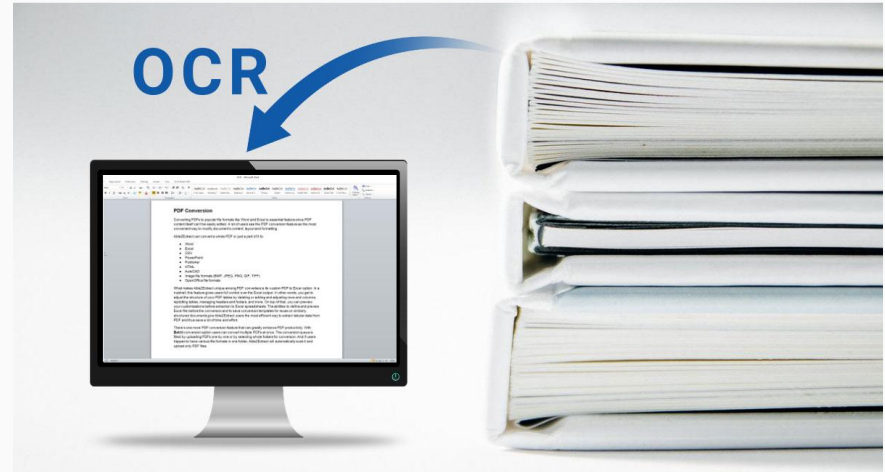
HONR 39900R Final Project

By Aiden Gonzalez



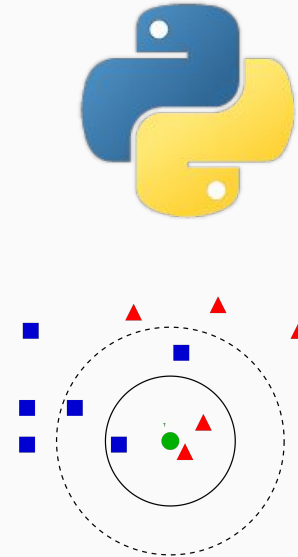
# What is Optical Character Recognition (OCR)?

Optical Character Recognition is the “conversion of typed, handwritten, or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo or from subtitle text superimposed on an image”



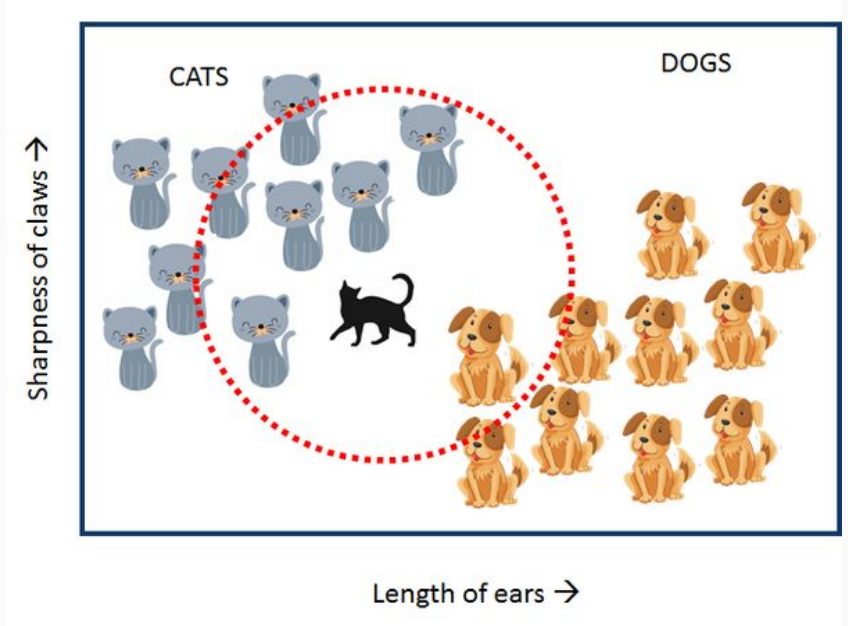
# What tools will we use to achieve this?

- Python: programming language with lots of machine learning libraries (aka modules, packages, toolkits etc) available
- OpenCV: Open-source computer vision library for Python that we'll use to "see" letters
- K-Nearest Neighbor Classification Algorithm (KNN): What we'll use to identify letters



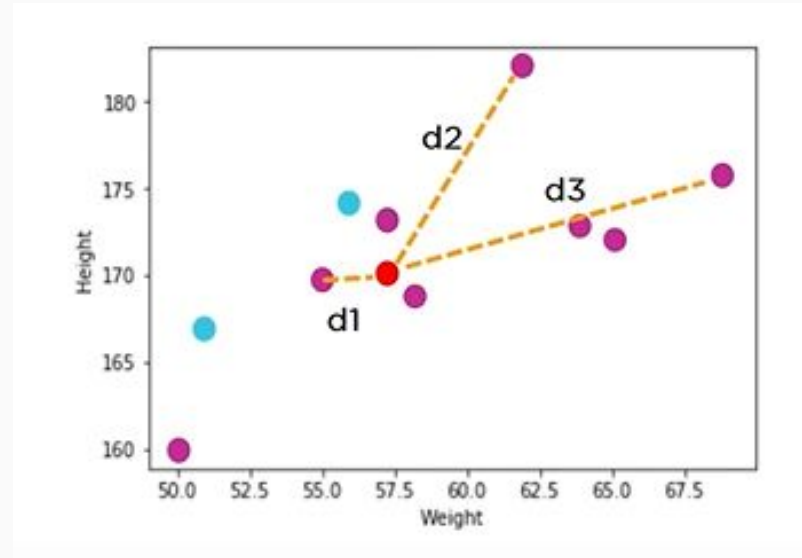
# How does KNN work?

- Start with a large classified dataset
- Input a datapoint
- Compare datapoint to find closest match in dataset
- Output matched classification



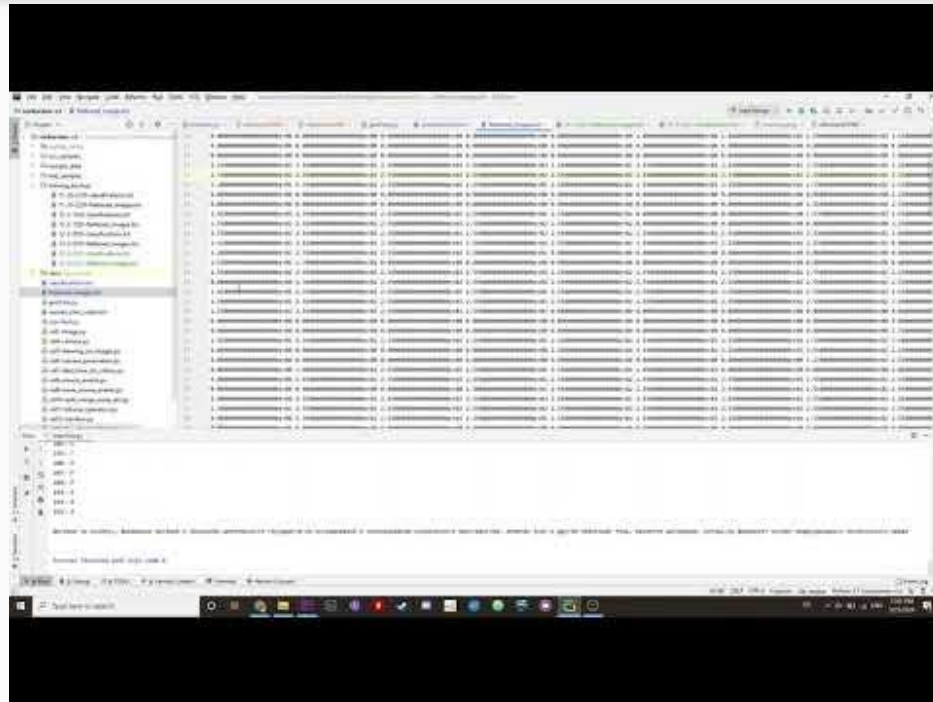
# KNN In-Depth

- Identify characteristics to localize inputted data
- Calculate distance between data point and all data points in the dataset
- Sort points by distance
- Select majority classification of the k nearest data points (k should be odd to eliminate “ties”)

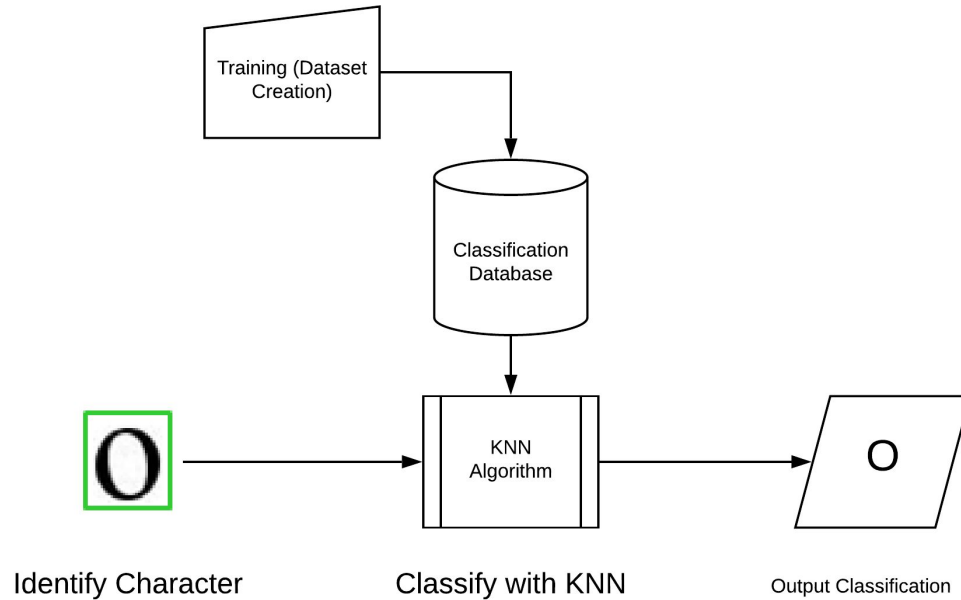




# Identifying Characters



# In Summary



# Sources

Wikipedia: Optical Character Recognition ([https://en.wikipedia.org/wiki/Optical\\_character\\_recognition](https://en.wikipedia.org/wiki/Optical_character_recognition))

KNN Algorithm - How KNN Algorithm Works With Example | Data Science For Beginners | Simplilearn  
(<https://www.youtube.com/watch?v=4HKqjENq9OU>)