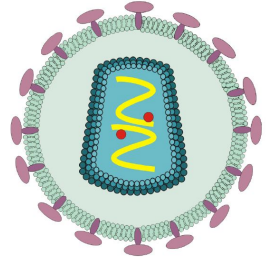
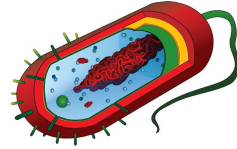
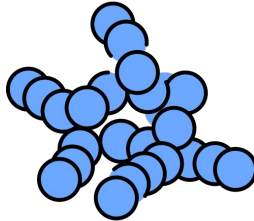
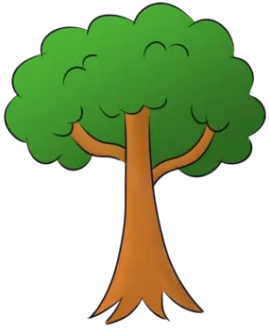


Codon Kingdom Classifier

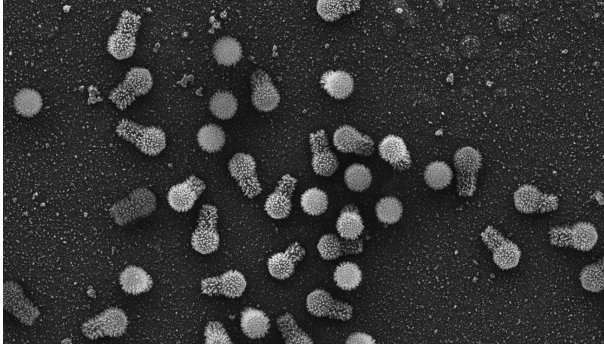
Sprint 3

Pier Bruno Pompili

Can you identify this organism by eye?



Now?



<https://www.pourlascience.fr/sd/microbiologie/des-virus-geants-tres-bacteriens-13161.php>



<https://sec-sem.blogspot.com/2011/01/bacteria-cultivation-mini-sem-image.html>

We can use DNA to differentiate, but

Can the usage of different codons be used to classify species by Kingdoms?

12 K Species

Phylogenetic information

64 Codons

Usage frequencies.

5 Kingdoms

Animalia, Plantae, Bacteria, Archaea, Virus

Objective

Attempt to classify codon usage in terms of lineage, by using machine learning methods to identify this genomics and evolutionary differences

Approach

Statistical Metrics for the ML models

Accuracy



Correct classification and labeling

Precision



Amount of variance and uncertainties of the data not explained

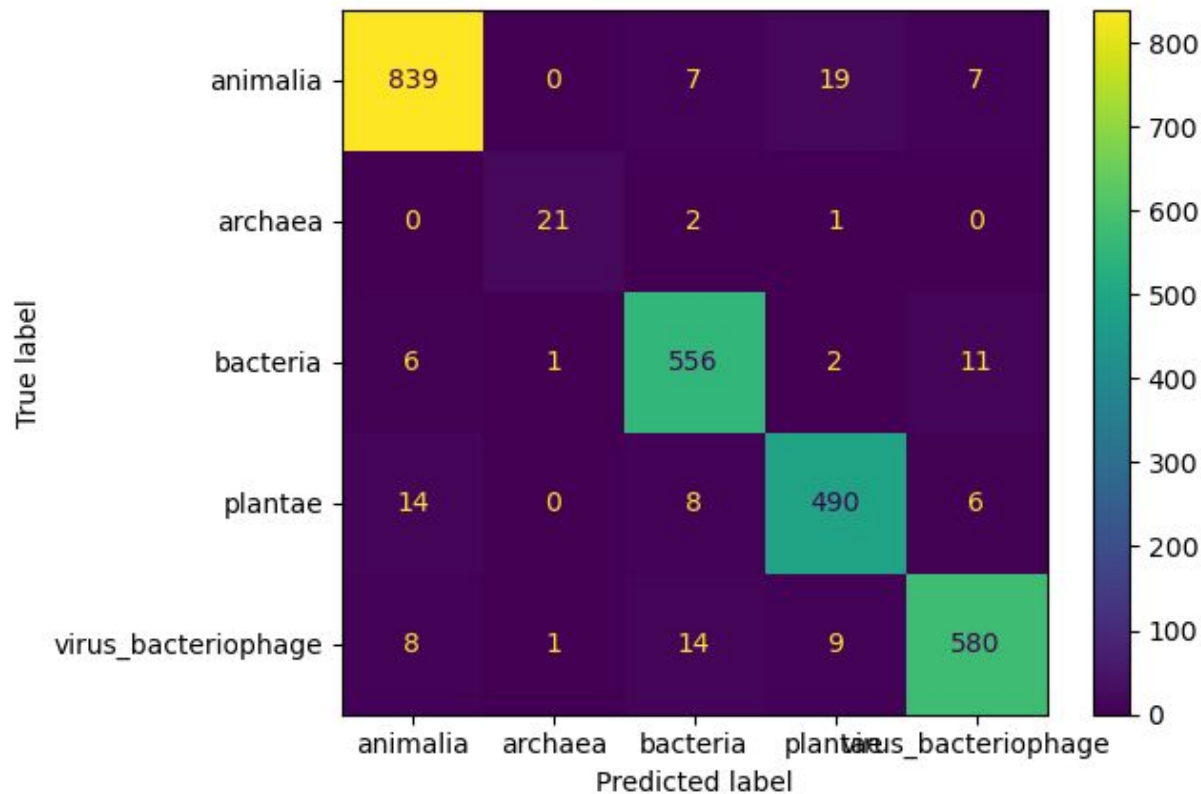
Recall



Sensitivity or True Positive Rate

Which model ?

KNN



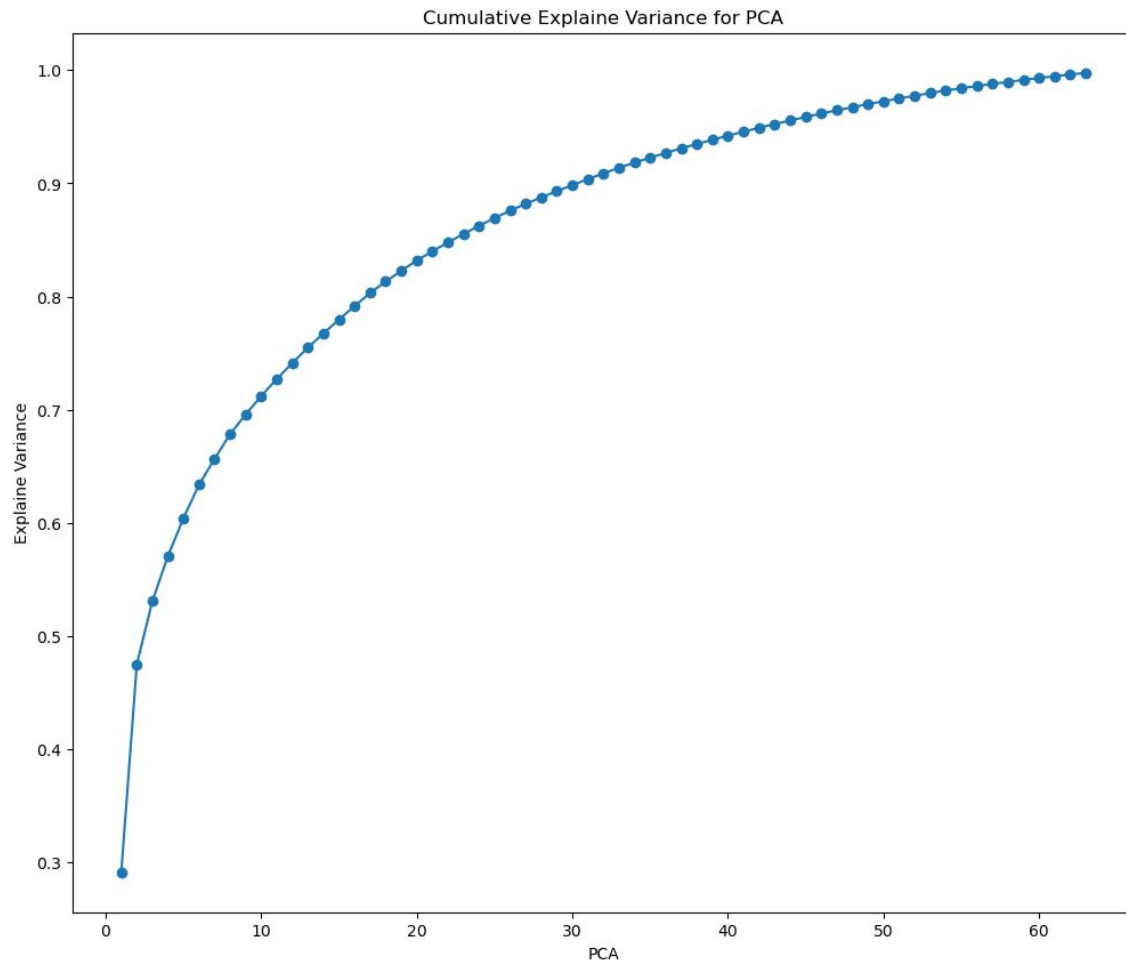
Accuracy = 96%

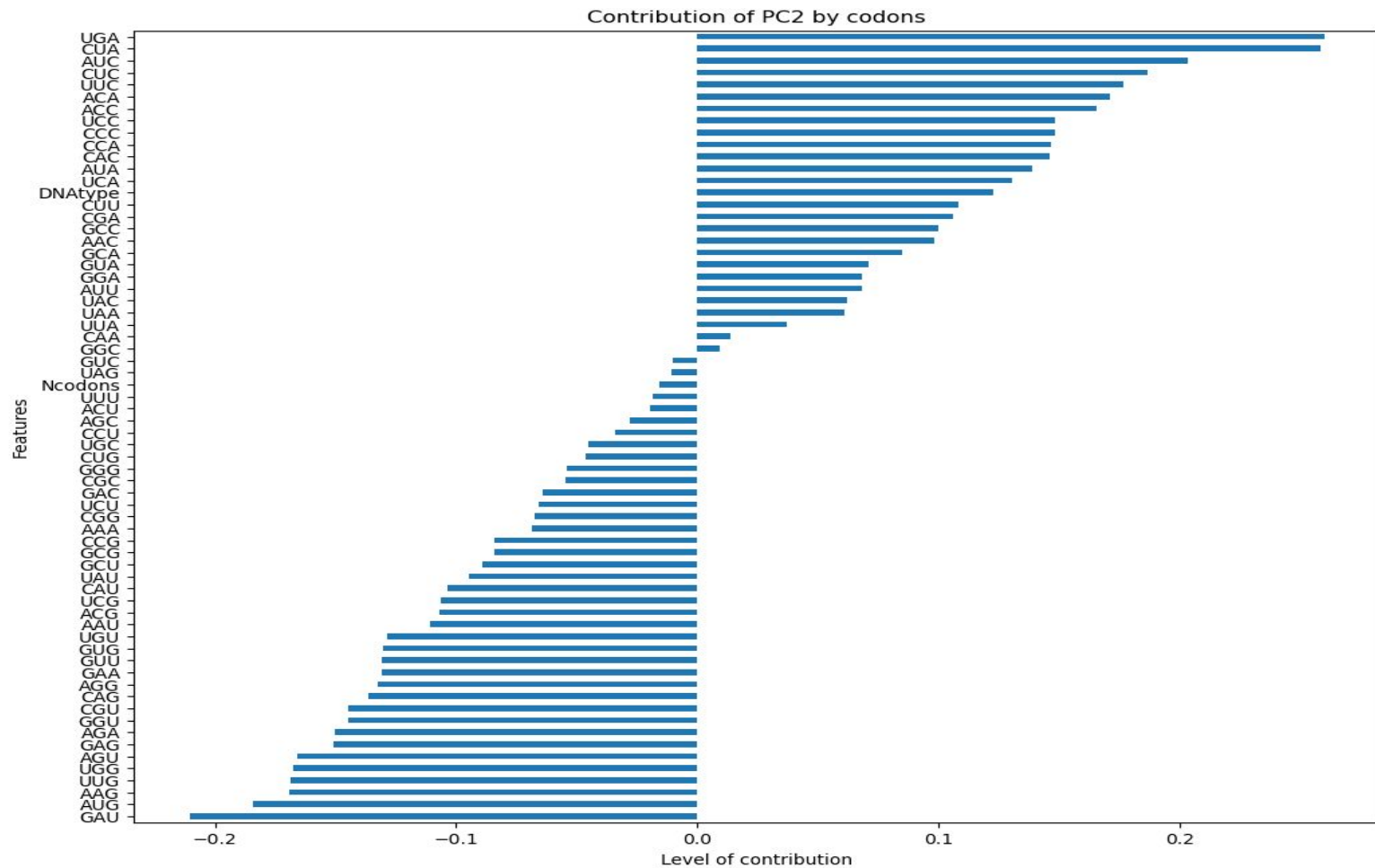
Precision = 95%

Recall = 94%

PCA Analysis

The principal components contributing to the classification task are **PC1, PC2** and **PC3**





Some Contributors

DNA type



Genomic, Mitochondrial and Chloroplast

UGA



Stop Codon

CUA



Leucine

Alternative source of carbon and nitrogen in energy-limited environments for Bacteria

Essential in regulating mammals metabolism

ACA



Threonine

Plant metabolism

Impact

Improving Taxonomic Classification

Refining kingdom definitions or to be used for the discovery of new species

Genetic Research

Facilitating the discovery of genetic markers that can be used for species identification

Bioinformatics

Machine Learning can answer biological questions for research



Thanks!