# E-Py-Genetics

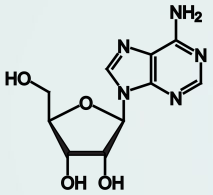
https://github.com/Karsten-Yan/pyberlin-ky

Dr. Karsten Yan

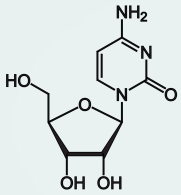
# TABLE OF CONTENTS


# Theory

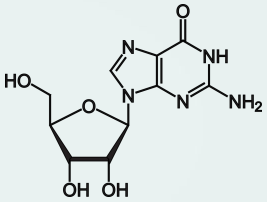
# Bases of the DNA



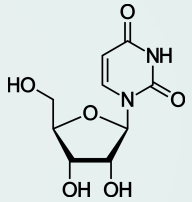

Adenosin




Cytidin

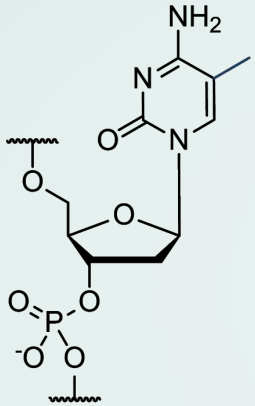



Guanosin

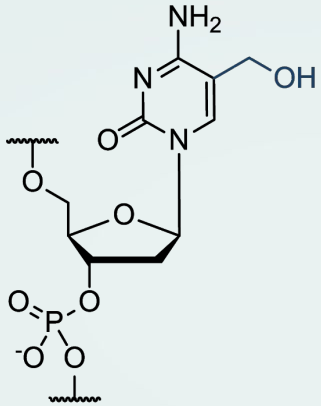



Thymidin

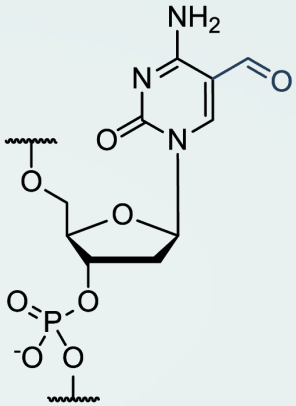
# Epigenetic DNA modifications



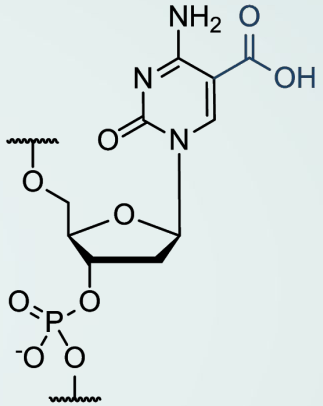
5-methyl
Cytidin



5-hydroxymethyl
Cytidin



5-formyl
Cytidin



5-carboxy
Cytidin

# Epigenetic DNA Modifications

01

Epigenetic modifications are relevant changes to the genome that do not alter the base sequence itself.

02

These modifications are, amongst other things, involved in the development of cancer and in evolutionary biology.

03

Currently established methods of detection of are comparatively slow and complex.

# Nanopore-Sequencing

# Data Basis

The utilized Dataset contains multiple reads of 5 different DNA strands of a length of 200 bases, which were either synthetically modified or left unmodified

# Notebook

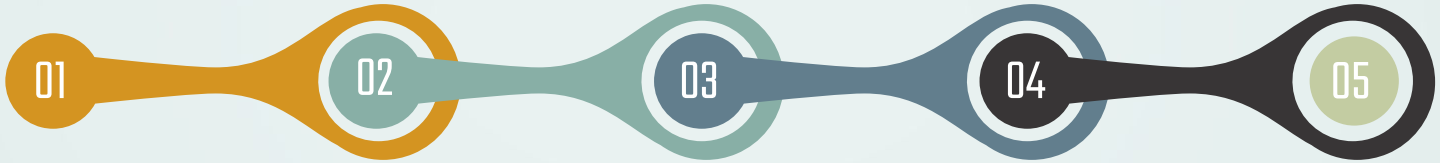

# Future Work

## Unsupervised Learning

Implementation of unsupervised outlier detection, to analyse unlabelled datasets

## Cloud Services

Port ML models to cloud services, to improve modelling times




## Neural Networks

Improving Recurrent Neural Network architecture

## Additional Datasets

Improve ML model using a bigger dataset

## Deployment

Package prediction models for application in real workflow


# Closing Remarks

# Thank you for Your attention!

- Thank you to neuefische, especially Larissa and Dirk, for the great Bootcamp, that allowed me to learn all the DataScience skills demonstrated in this capstone project
- And a big thank you to the DataScience cohort, who made it a pleasure to endure this Bootcamp with

https://www.linkedin.com/in/karsten-yan-854695151/

https://github.com/Karsten-Yan/ky-nf-capstone

# Questions?

# RESOURCES

## Dataset

- https://github.com/tleonardi/nanocompore/

## Opening Gif

- https://nanoporetech.com/how-it-works

# Tech Stack

- Python
- Pandas
- Scikit_learn
- Tensorflow
- Keras
- Ensemble Methods (XGBoost, ADABoost, Stacking)
- Matplotlib
- Seaborn

# THANKS

Do you have any questions?

youremail@freepik.com
+91  620 421 838
yourcompany.com