Protein Prediction I Exercise
Summer Term 2020
# (Preliminary) Project Outline

<table>
<thead>
<tr>
<th>Date</th>
<th>Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.04.20</td>
<td>Introduction</td>
</tr>
<tr>
<td>07.05.19</td>
<td>General information on SARS-CoV-2 and COVID-19</td>
</tr>
<tr>
<td>14.05.19</td>
<td>First results on literature research</td>
</tr>
<tr>
<td>21.05.19</td>
<td><em>No exercise (Christi Himmelfahrt)</em></td>
</tr>
<tr>
<td>28.05.19</td>
<td>Assessment of current literature</td>
</tr>
<tr>
<td>04.06.19</td>
<td>Distribution of protein features</td>
</tr>
<tr>
<td>11.06.19</td>
<td><em>No exercise (Fronleichnam)</em></td>
</tr>
<tr>
<td>18.06.19</td>
<td>Prediction of features for SARS-CoV-2 proteins</td>
</tr>
<tr>
<td>25.06.19</td>
<td>Refinement of results</td>
</tr>
<tr>
<td>02.07.19</td>
<td>09.07.19</td>
</tr>
</tbody>
</table>
Where we stand right now

- Every group chose their papers
- You checked the material on Machine Learning and answered the questions published on Piazza
- Any questions?
- Let’s discuss some “Machine Learning buzzwords”...
Machine Learning buzzwords

- Supervised vs. unsupervised learning
- Regression vs. classification
- Cross-validation
- Hyperparameter optimization
- Overfitting
- Redundancy reduction
- Class imbalance
- Performance assessment
- Overestimation of performance
Until the next exercise (in 2 weeks)

- Summarize the main methodology and results of your papers
- Prepare a “mini-review” of your papers
  - What is interesting/novel?
  - What needs more clarification?
  - What information is missing?
  - Are there any analyses missing?
  - Is the raw data accessible (to reproduce the results)?
  - ...
- Add your results to the rolling slide deck and present them in the next exercise
- Per group: 20 minutes (<15 minutes presentation, 5 minutes discussion)