**Minutes – Protein Disorder Dissertation**

14/10/22

* DNN demonstrated – One-hot encoding the protein – amino acid sequence and 15 element sliding window to make predictions about the disorder.
* PDB Database to train and validate the DNN.
* Looked at DisProt – database of disordered proteins – annotated disorders.
* Discussed former querying methods – REST and JSON. How FASTA was used.
* GraphQL to query the database – creates a graph – this is used for the input of the DNN.
* GraphQL works with the PDB Database.
* Clear understanding of what a DNN is now. It’s the more general term.
  + Fully connected – hidden layer that uses every neuron in the former layer as input to the next.
  + CNN – has multiple kernels.
  + RNN – weights shared – uses previous outputs as new inputs.
  + The above have their own subcategories too.
* Discussed agents – an agent could look like the following:
* Long term structure – to finalise ‘due’/prediction dates as goal tracking and show what can be feasibly done for when.
  + Create DNN – show benchmarking with different approaches and architectures.
  + Server for the agent.
    - Let users make a query.
    - To research: what can be displayed back to the user.
  + Stretch goal/future work: Create automatic refresh of agents learning requests.
    - Solves a problem where trained models do not make use of new data after they are finished.
* Goals for this week
  + PyTorch practise from the DL moodle. - learning about DNNs.
  + Look into what data is available on the PDB Database - to get comfortable interpreting it.
  + Review DISOPRED papers – useful as Background and will let me learn about different techniques.
  + Set up some goal trackers to help structure the project.