Technical Appendix
Hyper-parameters for DQN

Parameter	Value
Episodes	500
Hidden sizes	1st layer: 128, 2nd layer: 256
Learning rate	0.01
Discount rate γ	Variable, depending on the budget
Initial ϵ	1.0
ϵ decay rate	0.995
Replay buffer size	100,000

Table 1: Parameters used in the experiment section for DQN.

Parameter selection

For γ, we've tried [0.5, 0.6, 0.7, 0.75, 0.8, 0.9] and we have tuned this parameter based on the budget. We chose a gamma dependant on the budget because the budget directly determines the expected number of steps per episode.

Since the budget influences the number of steps per episode, we also tuned the number of steps required for the target model to update. We tested values in [25, 50, 100] number of steps and found that lower number of steps worked better for lower budget environments since these environments have shorter horizons.

We have also tried a larger network with hidden sizes 128, 258, 512, 1024, and 2048. We trained the network with 4,000 epochs, however, we did not find a significant difference in performance compared to the network with only two hidden layers.

Hyper-parameters for CBCTree

Parameter	Value
Min size clusters τ	10
cost scaler α	1

Table 2: Parameters used in the experiment section for CBC-Tree.

Dataset

Train/Validation Split

We randomly split the dataset into 80%/20% for the training and test splits, respectively. We reserved a small sample (10%) of the test set to create a validation set to tune the parameters, based on the total distance away from the closest points in the training set.

Computing Infrastructure

We used a AMD Ryzen 7 1700 CPU with 32 GB of RAM and a Nvidia RTX 3090 GPU. The average training time with our setup was 1 minute per 500 episodes with a dataset size of 3k samples. Tuning the model took us about 8 hours.