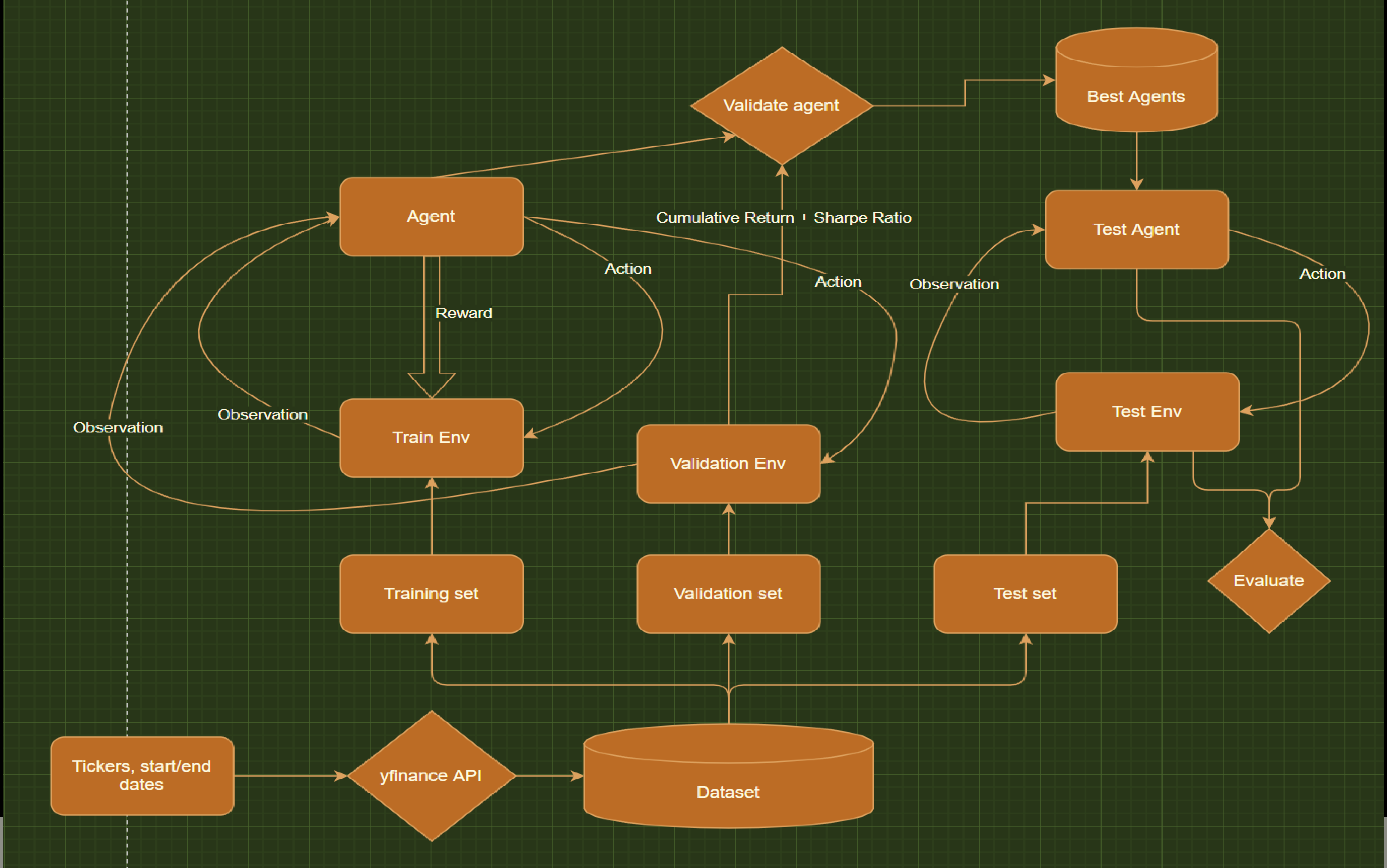
DEEP REINFORCEMENT LEARNING FOR OPTIMAL PORTFOLIO MANAGEMENT

Group Meeting 4

Summary of Models

- Training: 2008 – 17 (9 yrs), Validation: 2017 – 19 (2 yrs), Testing: 2019 – 21 (2 yrs)
- Observation space
  - *Instruments: AAPL, MSFT, V, JPM (+ JNJ, WMT)*
  - *Features: Open, High, Low, Close prices*
  - *Window: 50 trading days*
- Action Space
  - *DQN: Discretized holdings as multiples of 0.125, sum up to 1*
  - *DDPG: Continuous action space (0, 1)*
- Transaction costs: 0.25% of traded value
- Reward: Daily return / Daily return + Sharpe ratio


Validate agent
Best Agents
Agent
Cumulative Return + Sharpe Ratio
Test Agent
Action
Action
Observation
Action
Reward
Observation
Observation
Train Env
Validation Env
Test Env
Training set
Validation set
Test set
Evaluate
Tickers, start/end dates
yfinance API
Dataset

Hyperparameters

	DQN	DDPG
Discount Factor (gamma)	~0.5	~0.8
Learning Rate	~0.00005	~0.00001
Batch size	64/128	64/128
Update frequency	1 step	100 steps
Target update	1000	NA
Polyak update (tau)	1	0.005
Replay Buffer size	10000	10000

Exploration

- DQN
  - *epsilon-greedy strategy*
    - epsilon: 1 – 0.6
    - epsilon decay: 0.4 – 0.6
    - epsilon min: 0.01
- DDPG
  - *Ornstein Uhlenbeck Action Noise*
  - *Normal Action Noise*

26th July – 16th August: Individual Meeting 2

- Expected: 90% of report writing completed
- Integration of code base

FIN