PPO Market Bot

Mathys Vinatier

Seoul National University

October 1, 2025

Actor-Critic Layouts

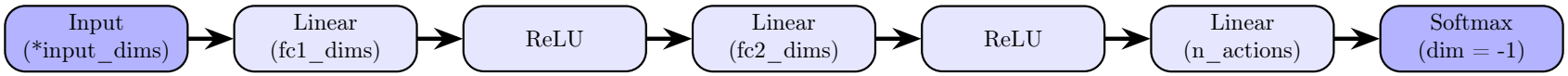

Figure 1: Actor network architecture (Softmax output)

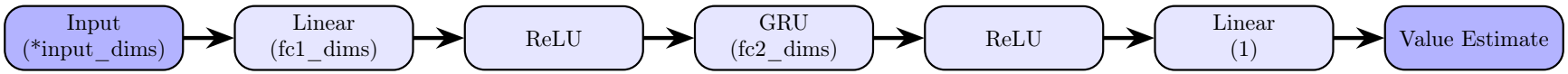

Figure 2: Critic network architecture (GRU layout)

Training the Agent

Algorithm 1: PPO Agent Training Loop

Input: env, n_games, N, $batch\_size$, α, n_epochs
Output: Trained PPO agent, episode rewards
Initialize PPO_agent;
for $i \leftarrow 1$ **to** n_games
    $s \leftarrow env.reset()$;
    $done \leftarrow$ False;
    $score \leftarrow 0$;
    while $done = False$
        $A_{valid} \leftarrow env.get\_valid\_actions()$;
        $(a, \text{probs}, v) \leftarrow agent.choose\_action(s, A_{valid})$;
        $(s', r, done, info) \leftarrow env.step(a)$;
        $score \leftarrow score + r$;
        $n\_steps \leftarrow n\_steps + 1$;
        $agent.remember(s, a, \text{probs}, v, r, done)$;
        if $n\_steps \bmod N = 0$ **then**
            $agent.learn()$;
            $learn\_iters \leftarrow learn\_iters + 1$;
return trained agent;

First Result after 500 episodes training

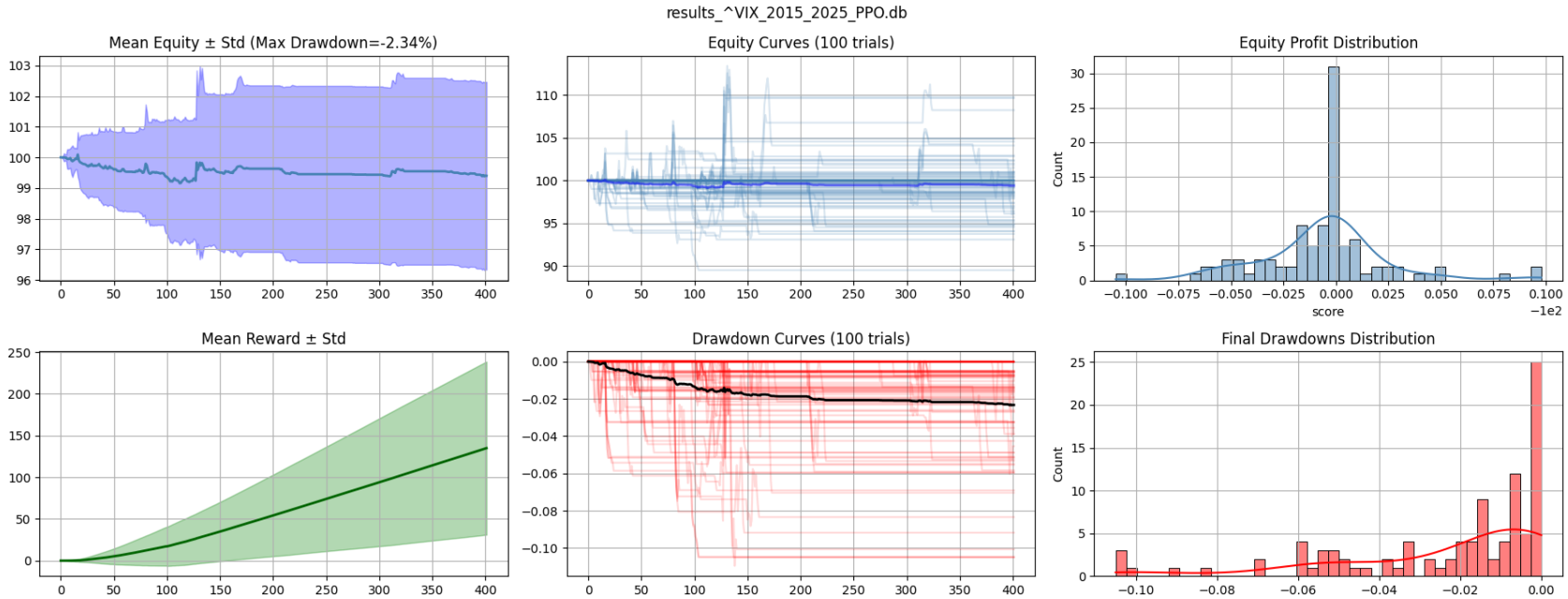

Figure 3: First ouput for PPO agent

Trade Station Server

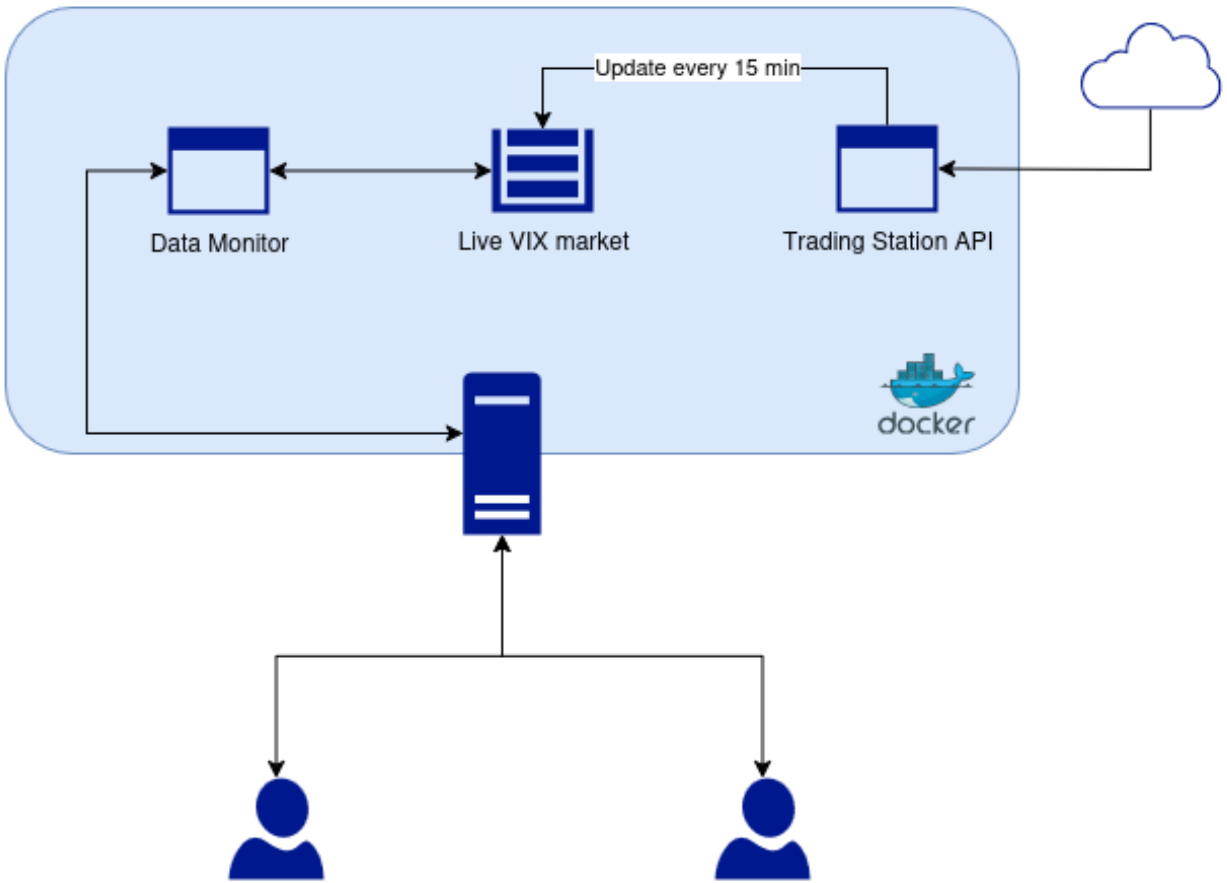

Figure 4: Architecture of the Service

Trade Station Server

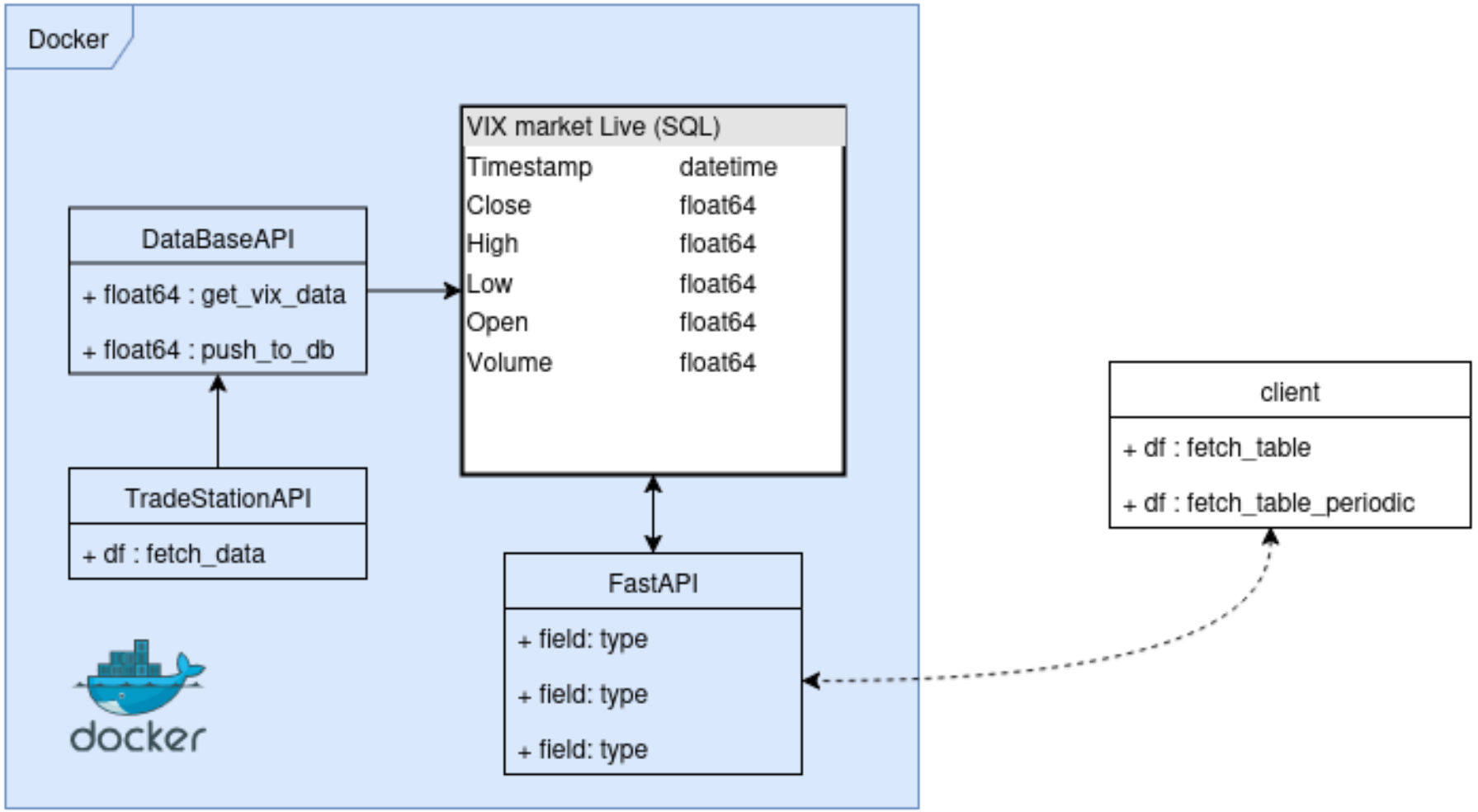

Figure 5: UML of the Service

Thank You