Reinforcement Learning

For reinforcement learning we used the OHLCV data from binance API to stream live data to our Actor-Critic model. Live data was collected every minute, where the date is an object type and the OHLCV is float64.The Actor-Critic model was developed using the Tensorflow API, where we used CNN, LSTM and Dense layers for them both. Our action states for our agent were Buy, Sell and Hold and our reward algorithm for our agent is based off of TD(temporal difference)learning. We created our environment with parameters that use data from past interactions to make decisions like our variables order history and market history.

The agent interacts with the environment by using lookbacks of previous OHLCV data and the conditions of its wallet, balance or money available to buy coins, networth and the value of balance and current coins, how much it bought, sold or held.

The environment calculates the reward for the agent based on the total amount of coins and at what price it bought those coins and what they are worth in the current time step. Typical indicators like simple moving average with increasing window sizes as well Bollinger Bands, Parabolic Stop and Reverse, MACD and RSI are used. We approached the problem by adding technical indicators incrementally, and recording the improvement in the agent capabilities. Some examples of what the agent is doing by using different indicators are below.
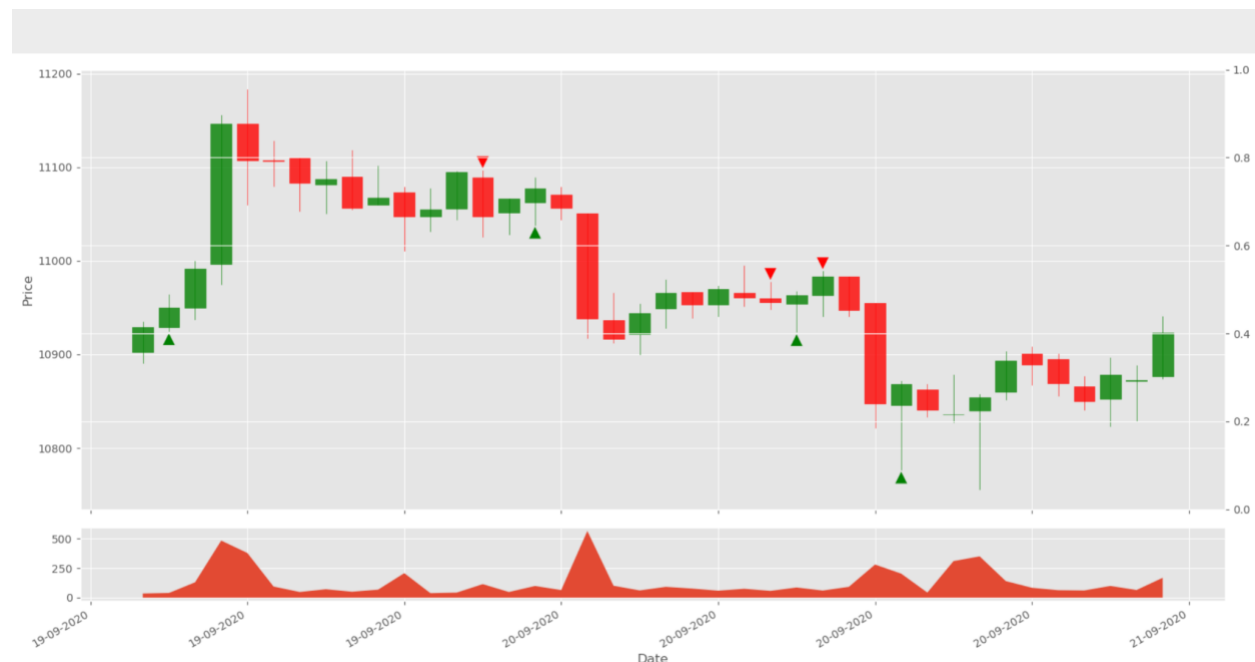
(Candlestick chart with BUY (GREEN) / SELL (RED) signals without reward policy)

(Candlestick chart with BUY (GREEN) / SELL (RED) signals and the WALLET BALANCE of the agent (BLUE))
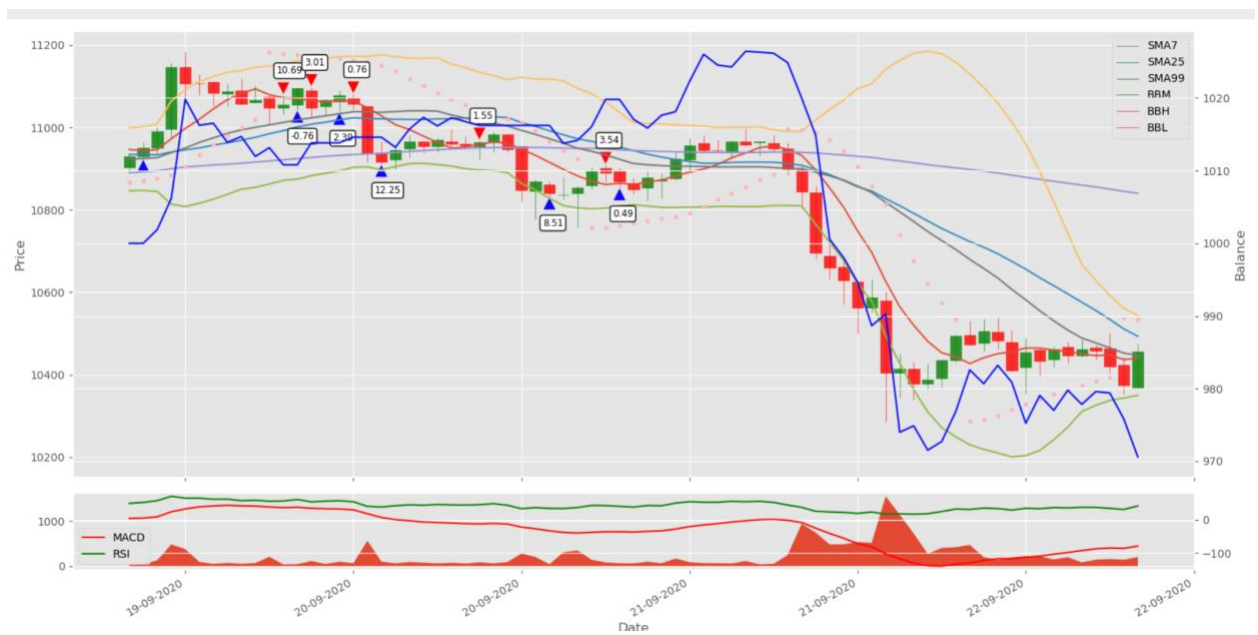
(Candlestick chart with BUY (GREEN) / SELL (RED) signals + WALLET BALANCE of the agent (BLUE) + Technical Indicators + Agent Reward (in boxes))