Algorithm MAB Epsilon Greedy

```
Initialize, for $a = 1$ to $k$:
$Q(a) \leftarrow 0$
$N(a) \leftarrow 0$

for $t$ in range($len(data)$) do
    $A_t \leftarrow \begin{cases} \text{a random action with probability } \epsilon \\ \text{argmax}_a\, Q(a) \text{ with probability } 1 - \epsilon \end{cases}$
    $R_t \leftarrow \text{bandit}(A_t)$
    $N(A_t) \leftarrow N(A_t) + 1$
    $Q(A_t) \leftarrow Q(A_t) + \frac{1}{N(A_t)}[R_t - Q(A_t)]$
end for
```