



دانشگاه شهید بهشتی  
دانشکده مهندسی و علوم کامپیوتر

استفاده از یادگیری تقویتی در صندوق‌های سرمایه‌گذاری رمزارزهای انرژی‌های سبز

پروژه کارشناسی

پارسا نوری

دکتر مائده مشرف دهکردی

زمستان ۱۴۰۲

## چکیده

بازار اعتبارات کربن پس از توافق پاریس به یکی از بازارهای مهم جهانی تبدیل شده است. اعتبارات کربن به عنوان رمزارزها نیز می‌توانند برای افزایش شفافیت در این بازار مورد استفاده قرار گیرند. یکی از راهکارهای مهمی که سیستم‌های اعتبار کربن باید ارائه دهند، این است که اعتبار رمزارز به ارزی قابل معامله تبدیل شود. راهکار ارائه شده در اینجا استفاده از استخرهایی است که از یادگیری تقویتی بهره می‌برند. دلیل استفاده از این استخرها، توازن در فرکانس تولید رمزارز خروجی است. برای حل این مشکل، از شبکه یادگیرنده تقویتی مبتنی بر شبکه عمیق Q استفاده کردیم. اعتبارات کربن ممکن است ارزش‌های مختلفی داشته باشند، به توجه به نوع ارز موجود در آن‌ها، و راهکاری ارائه شده است تا این تفاوت‌ها در این استخرها مدیریت شوند. نحوه استفاده از شبکه یادگیرنده عمیق در قراردادهای هوشمند نیاز به حل چالش‌های مختلف داشته و ما به آن‌ها پرداختیم. در پایان، یک راهکار برای آماده‌سازی این یادگیرنده برای استفاده در دنیای واقعی ارائه شده است که شامل شبیه‌سازی محیطی است که عامل در آن فعالیت می‌کند تا عملکرد آن مورد سنجش قرار گیرد.

## فهرست مطالب

1	فصل اول: کلیات	1
2	1-1 مقدمه	2
4	2-1 برخی مفاهیم	4
4	1-2-1 مفهوم ERC-20	4
4	1-2-2 مفهوم ERC-721	4
5	3-2-1 مفهوم ERC-1155	5
6	3-1 بیان مسئله	6
6	4-1 کلیات روش پیشنهادی	6
7	5-1 ساختار پروژه	7
8	فصل دوم: سیستم‌های رمزارزی اعتبار کرین موجود	8
9	1-2 مقدمه	9
9	2-2 بررسی سیستم توکان	9
9	1-2-2 بازار انرژی توکان	9
11	2-2-2 سیستم پل در توکان	11
13	3-2-2 بررسی استخرها در سیستم توکان	13
14	2-3 بررسی سیستم انرژی وب اریجین	14
15	1-3-2 بررسی معماری انرژی وب اریجین	15
19	4-2 جمع‌بندی	19
21	فصل سوم: بررسی روش‌های یادگیری تقویتی	21
22	1-3 مقدمه	22

22	بررسی روش‌های مختلف یادگیری تقویتی	2-3
22	تقویت یادگیری بدون مدل	1-2-3
26	یادگیری تقویتی با مدل	3-2-2
27	یادگیری تقلیدی	3-2-3
28	یادگیری تقویتی عمیق	4-2-3
29	استراتژی‌های کاوش	3-3
30	اپسیلون-حریص	1-3-3
31	مرز بالای اعتماد	2-3-3
31	نمونه‌برداری تامپسون	3-3-3
32	جمع‌بندی	4-3
34	فصل چهارم روش پیشنهادی و روش پیاده‌سازی	
35	مقدمه	1-4
35	نحوه پیاده‌سازی بخش تبدیل اعتبار کرن به رمزارز	2-4
36	توکن‌های ERC1155 با یکدیگر متفاوت هستند	4-3
36	راهکار یادگیری تقویتی برای تبدیل ERC-1155 به ERC-20	4-4
38	چرا یادگیری تقویتی؟	5-4
39	انتخاب مدل یادگیرنده تقویتی	6-4
39	یادگیرنده تقویتی شبکه یادگیری عمیق Q	4-6-1
40	یادگیرنده عمیق SARSA	4-6-2
41	تعیین تابع پاداش	4-7
41	تابع قیمت	4-7-1
41	تابع سرمایه‌گذاری	4-7-2
41	مصالحه کاوش و بهره‌برداری	4-8
43	معماری شبکه عصبی	4-9

44 .....	شبیه‌سازی.....	4-10
41 .....	تست .....	4-11
47 .....	نحوه پیاده‌سازی .....	4-12
49 .....	پیاده‌سازی قراردادهای هوشمند .....	4-12-1
50 .....	پیاده‌سازی سرور backend .....	4-12-2
50 .....	پیاده‌سازی یادگیری تقویتی .....	4-12-3
51 .....	جمع‌بندی .....	4-13

## فهرست شکل‌ها

- تصویر 1 مکانیزم تبدیل ارز غیر قابل عوض به قابل عوض ..... 7
- تصویر 2 شرح پل توکان ..... 11
- تصویر 3 - معماری انرژی وب اریجین ..... 16
- تصویر 4 - فرایند سرمایه گذاری ..... 35
- تصویر 5- توزیع بتا منبع وبسایت Towards Data Science ..... 43
- تصویر 6 - معماری پیاده‌سازی ..... 47

## فهرست جدول‌ها

جدول 1- مقایسه توکان و انرژی وب اریجین.....	19
جدول 2- عوامل موثر بر میزان تولید رمزارز خروجی استخر.....	38
جدول 3- احتمالات تغییر فرکانس سرمایه‌گذاری و قیمت حین شبیه‌سازی.....	46
جدول 4- امتیازهای عامل‌های مختلف.....	46
جدول 5 - زمان یادگیری عامل‌ها مختلف.....	47
جدول 6- جدول امتیاز دهی به استخرهای اعتبارات کربن رمزارزی.....	48
جدول 7 - جدول کدگذاری اطلاعات رمزارزهای ERC1155.....	49





## فصل اول: کلیّات

## 1-1 مقدمه

کشورهای توافق پاریس، یا موافقت‌نامه پاریس در خصوص تغییر اقلیم، یک توافق بین المللی است که در ۱۲ دسامبر ۲۰۱۵ در پاریس، فرانسه، امضاء شد. این توافق به عنوان قطبی‌ترین و کامل‌ترین توافق در زمینه تغییر اقلیم تاکنون شناخته می‌شود. اهداف اصلی توافق پاریس عبارتند از کاهش افزایش دما به زیر ۲ درجه سانتی‌گراد نسبت به دوران قبل از صنعت‌گری و تلاش برای محدود کردن این افزایش به حداکثر ۱.۵ درجه سانتی‌گراد. توافق‌نامه پاریس تعهدات مالی، تعهدات کاهش گازهای گلخانه‌ای، و توسعه پایدار را نیز در بر می‌گیرد. این توافق همچنین بر اهمیت همکاری جهانی در مقابله با تغییر اقلیم تأکید دارد و تعهد به ارائه حمایت‌های مالی و فنی به کشورهای در حال توسعه را دارد. [1]

اعتبار کربن یک واحد اندازه‌گیری است که کاهش یا حذف یک تن دی‌اکسید کربن یا سایر گازهای گلخانه‌ای از جو را نشان می‌دهد. اعتبارات کربن معمولاً توسط پروژه‌هایی ایجاد می‌شوند که به کاهش انتشار گازهای گلخانه‌ای کمک می‌کنند. منابع اعتبار کربن می‌توانند شامل موارد زیر باشند:

- استفاده از انرژی‌های تجدیدپذیر، مانند انرژی خورشیدی و باد
- بهبود بهرموری انرژی، مانند عایق‌بندی ساختمان‌ها
- کاشت درخت، که دی‌اکسید کربن را از جو جذب می‌کند
- کاهش انتشار گازهای گلخانه‌ای از صنایع، مانند کاهش استفاده از سوخت‌های فسیلی

اعتبارات کربن می‌توانند به روش‌های مختلفی استفاده شوند. یکی از روش‌های استفاده از آنها، خرید و فروش در بازارهای کربن است. در این بازارها، شرکت‌ها و سازمان‌هایی که انتشار گازهای گلخانه‌ای خود را کاهش نمی‌دهند، می‌توانند اعتبارات کربن را از شرکت‌ها و سازمان‌هایی که انتشار گازهای گلخانه‌ای خود را کاهش داده‌اند، خریداری کنند. این امر به شرکت‌ها و سازمان‌ها کمک می‌کند تا به اهداف کاهش انتشار گازهای گلخانه‌ای خود دست یابند.

اعتبارات کربن همچنین می‌توانند برای جبران انتشار گازهای گلخانه‌ای استفاده شوند. به عنوان مثال، یک شرکت هواپیمایی ممکن است اعتبارات کربن را خریداری کند تا انتشار گازهای گلخانه‌ای ناشی از پروازهای خود را جبران کند.

اعتبارات کربن یک ابزار مهم برای مقابله با تغییر اقلیم هستند. آنها می‌توانند به شرکت‌ها و سازمان‌ها کمک کنند تا به اهداف کاهش انتشار گازهای گلخانه‌ای خود دست یابند و همچنین می‌توانند به جبران انتشار گازهای گلخانه‌ای کمک کنند. [2]

رمزارز اعتبار کربن<sup>1</sup> یک نوع رمزارز است که با اعتبارات کربن پشتیبانی می‌شود. این اعتبارات کربن نماینده یک تن دی‌اکسید کربن هستند که از جو حذف شده است. این اعتبارات از طریق پروژه‌هایی که انتشار گازهای گلخانه‌ای را کاهش می‌دهند، ایجاد می‌شوند. سپس توسعه‌دهندگان پروژه می‌توانند این اعتبارات را به افراد یا شرکت‌هایی که می‌خواهند اثر محیطی خود را جبران کنند، بفروشند. برخی از شرکت‌های فناوری، اعتبارات کربن و توکن‌های<sup>2</sup> رمزارز را با هم ترکیب کرده‌اند. به طور کلی، رمزارزهای اعتبار کربن می‌توانند باعث شوند تا بازار اعتبارات کربن شفاف‌تر و قابل دسترسی‌تر شود. [3]

بازار اعتبار کربن یک بازار جهانی است که در آن شرکت‌ها و افراد می‌توانند برای جبران انتشار کربن خود، اعتبارات کربن خریداری کنند. این اعتبارات از پروژه‌هایی مانند کاشت درختان، توسعه انرژی تجدیدپذیر و کاهش مصرف انرژی به دست می‌آیند. بازار اعتبار کربن سنتی دارای برخی از مشکلات است، از جمله:

- سوءاستفاده: برخی از شرکت‌ها از اعتبارات کربن برای دستیابی به اهداف زیست‌محیطی خود بدون انجام اقدامات واقعی برای کاهش انتشار کربن استفاده می‌کنند.
- عدم شفافیت: ردیابی اعتبارات کربن در بازار سنتی دشوار است.
- هزینه بالای معاملات: معاملات در بازار سنتی می‌تواند گران باشد.

بازار اعتبار کربن در رمز ارزها یک بازار جدید است که در آن اعتبارات کربن به عنوان توکن‌های رمزنگاری شده معامله می‌شوند. این بازار پتانسیل آن را دارد تا برخی از مشکلات بازار سنتی اعتبار کربن را برطرف کند. از مزایای این بازار می‌توان گفت:

- شفافیت: توکن‌های رمزنگاری شده اعتبار کربن را می‌توان به راحتی ردیابی کرد.
- کارایی: معاملات در بازار رمز ارزها معمولاً سریع‌تر و کارآمدتر از معاملات در بازار سنتی است.
- دسترسی: بازار رمز ارزها دسترسی بیشتری به سرمایه‌گذاران و شرکت‌ها دارد.

#### چالش‌ها

مقررات: مقررات بازار اعتبار کربن در رمز ارزها هنوز در حال توسعه است.

سوءاستفاده: توکن‌های رمزنگاری شده اعتبار کربن نیز می‌توانند مورد سوءاستفاده قرار گیرند.

---

<sup>1</sup> Carbon Credit Cryptocurrency

<sup>2</sup> Token

بازار اعتبار کربن در رمز ارزها هنوز در مراحل اولیه توسعه خود قرار دارد. با این حال، این بازار پتانسیل آن را دارد تا به مقابله با تغییر اقلیم کمک کند.

بازار اعتبار کربن در رمز ارزها یک بازار نوظهور است که پتانسیل آن را دارد تا به مقابله با تغییر اقلیم کمک کند. این بازار هنوز در مراحل اولیه توسعه خود قرار دارد و چالش‌هایی نیز در پیش دارد. با این حال، این بازار پتانسیل آن را دارد تا بازار اعتبار کربن سنتی را متحول کند. [4]

اعتبارات کربن در سیستم‌های رمزارزهای انرژی‌های سبز به طور عمومی به صورت توکن‌های غیرقابل معاوضه در سیستم رمزارزهای تعریف می‌شود. یکی از راهکارهایی که سیستم‌های رمزارزی بازار انرژی بایستی ارائه کنند مکانیزم تبدیل این توکن‌ها به توکن‌های قابل معاوضه با استاندارد است تا از آن‌ها بتوان به عنوان یک ارز استفاده کرد..

## 2-1 برخی مفاهیم

### 1-2-1 مفهوم ERC-20

ERC-20 یک استاندارد قرارداد هوش مصنوعی در بلاکچین Ethereum است که برای ایجاد توکن‌های قابل تبادل بر اساس بلاکچین Ethereum استفاده می‌شود.

ERC-20 تعریف مشخصی برای توکن‌های مبادله‌پذیر یا تعویض‌پذیر ایجاد می‌کند و مشخصاتی مانند نام توکن، نماد توکن، تعداد کل توکن‌ها، تعداد اعشار توکن، و قابلیت‌های انتقال توکن را تعریف می‌کند. این استاندارد امکان تبادل و معامله توکن‌ها بین مخاطبان مختلف را فراهم می‌کند و به توکن‌های ERC-20 اجازه می‌دهد تا به راحتی در کیف‌های مختلف Ethereum ذخیره و مدیریت شوند.

به عبارت دیگر، ERC-20 به توکن‌های مبادله‌پذیر بر اساس بلاکچین Ethereum یک ساختار و استاندارد مشخص می‌دهد تا امکان انتقال، ذخیره‌سازی و معامله‌ی آنها در اکوسیستم Ethereum را بهبود بخشد و این امکان را به کاربران و توسعه‌دهندگان می‌دهد که به راحتی با توکن‌های مختلف ارتباط برقرار کنند.

### 1-2-2 مفهوم ERC-721

ERC-721 یک استاندارد قرارداد هوش مصنوعی برای توکن‌های غیرقابل تعویض<sup>3</sup> در بلاکچین Ethereum است. توکن‌های ERC-721 یک منحصر به فرد هستند و هر یک دارای ویژگی‌ها و مشخصات

---

<sup>3</sup> Non-Fungible Tokens

منحصر به فردی هستند که آنها را از یکدیگر متمایز می‌کنند. به عبارت دیگر، هر توکن ERC-721 نماینده یک دارایی یا عنصر خاص در دنیای واقعی یا مجازی است.

از آنجا که توکن‌های ERC-721 منحصر به فرد هستند و هر یک دارایی یا عنصر خاصی را نمایندگی می‌کنند، این استاندارد به طور گسترده‌ای در برنامه‌ها و بازی‌های آنلاین به کار می‌رود که نیاز به تبادل دارایی‌های منحصر به فرد دارند.

## ERC-1155 مفهوم 3-2-1

ERC-1155 یک استاندارد قرارداد هوش مصنوعی برای توکن‌های چندپراکنشی<sup>4</sup> در بلاکچین Ethereum است. ERC-1155 به توسعه‌دهندگان امکان ایجاد توکن‌هایی با انواع مختلفی از دارایی‌ها را می‌دهد. به عبارت دیگر با ERC-1155، یک قرارداد هوش مصنوعی می‌تواند توکن‌هایی ایجاد کند که همزمان انواع مختلفی از دارایی‌ها را نمایندگی کنند.

ویژگی‌های مهم-ERC 1155 شامل موارد زیر هستند:

1. چندپراکنشی: توکن‌های-ERC 1155 می‌توانند انواع مختلفی از دارایی‌ها را نمایندگی کنند. این انواع می‌توانند شامل توکن‌های غیرقابل تعویض (NFTs) و توکن‌های قابل تعویض (FTs) باشند. این امکان به توسعه‌دهندگان اجازه می‌دهد تا توکن‌هایی با ویژگی‌ها و قابلیت‌های مختلف ایجاد کنند.
2. کارایی بهتر-ERC: 1155 باعث کاهش هزینه‌ها و پیچیدگی‌ها در مقایسه با استفاده از استانداردهای جداگانه برای NFTs و FTs می‌شود. این به معنای کاهش حجم داده‌های ذخیره‌سازی و تراکنش‌ها در بلاکچین است.
3. انتقال متعدد: توکن‌های-ERC 1155 امکان انتقال متعدد را به صورت یک تراکنش در بلاکچین فراهم می‌کنند. این بدین معناست که می‌توانید تعداد زیادی از توکن‌ها را به یک مقصد انتقال دهید.
4. توسعه‌پذیری-ERC: 1155 امکان توسعه‌دهندگان را در ایجاد برنامه‌ها و بازی‌هایی که نیاز به تعامل با توکن‌های متنوع دارند، تسهیل می‌کند.

---

<sup>4</sup> Multi-Fungible Tokens (MFT)

ERC-1155 به عنوان یک استاندارد انعطاف‌پذیر و کارا در برنامه‌نویسی توکن‌های بلاکچین و ایجاد بازی‌های مبتنی بر بلاکچین با انواع مختلفی از دارایی‌ها بسیار مورد توجه قرار گرفته است.

## 1-3 بیان مسئله

سیستم بازار اعتبارهای کریب یک مکانیزم برای تبدیل اعتبار کریب به ارز قابل معاوضه نیاز دارند. این سیستم تا به حال به این صورت بوده که به ازای هر تن کریب که از کره زمین حذف شده است یا خواهد شد، یک واحد رمزارز قابل عوض به فرد ارائه دهنده اعتبار داده می‌شود.

این کار اگرچه می‌تواند مناسب باشد ولی تنها راهکار نیست. داشتن یک بازار تبدیل توکن‌های بلاعوض به توکن‌های قابل تعویض می‌تواند با تعویض مکانیزم گفته شده با یک استخر سرمایه‌گذاری رمزارزی بسیار تغییر کند. به این صورت که به افراد دارنده اعتبار کریب، به ازای مدت زمان نگهداری اعتبار نام‌برده در استخر با فرکانسی توکن قابل تعویض ارائه شود.

حال سوالی که مطرح می‌شود آن است که فرکانس ارائه توکن قابل تعویض در ازای توکن غیرقابل تعویض به چه صورت باشد؟

## 1-4 کلیات روش پیشنهادی

در این جا می‌توان از یادگیری تقویتی<sup>5</sup> بهره برد به شرط آن که بتوان مولفه‌های آن را به درستی تعریف نمود. این مولفه‌ها به صورت زیر هستند:

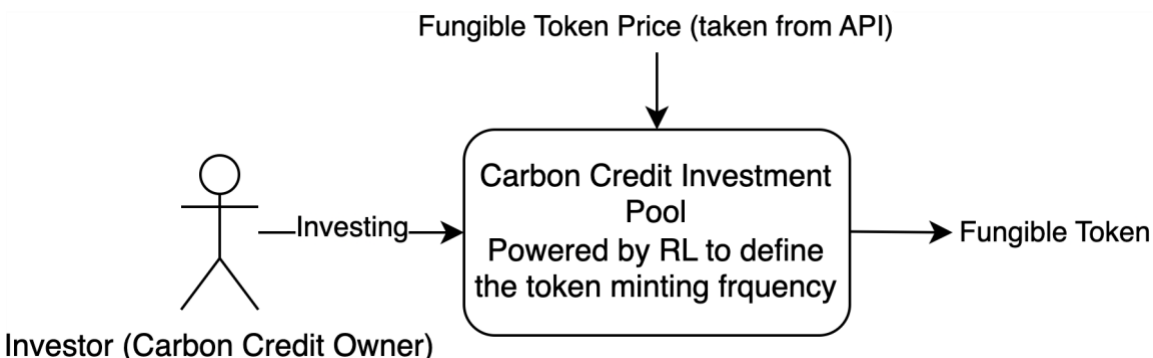
1. **وضعیت:** وضعیت در این مسئله قیمت ارز قابل تعویض و میزان فرکانس سرمایه‌گذاری در استخر یا صندوق سرمایه‌گذاری می‌باشد.
2. **امتیاز و پاداش:** امتیاز در این مسئله افزایش میزان سرمایه‌گذاری یا افزایش نرخ ارز خروجی است. همچنین کاهش این دو نیز می‌تواند عامل منفی در نظر گرفته شود.

---

<sup>5</sup> Reinforcement Learning

3. **عمل عامل:** عامل می‌تواند با تغییر نرخ تولید ارز قابل معاوضه میزان موجودیت آن را در بازار تغییر بدهد.

4. **تاثیر عامل بر محیط و بیان تعادل لازم:** عامل یادگیرنده تقویتی در اینجا با تعیین میزان موجودیت ارز در بازار می‌تواند روی قیمت تاثیر گذار باشد. افزایش قیمت می‌تواند منجر به افزایش میزان سرمایه‌گذاری نیز شود. پس کاهش نرخ توضییه منجر به افزایش قیمت و افزایش قیمت نیز منجر به افزایش میزان



**Investor (Carbon Credit Owner)**

سرمایه‌گذاری جهت کسب ارز قابل معاوضه می‌شود. از طرف دیگر هرچه میزان تولید ارز قابل تعویض بیشتر باشد میزان این صندوق سرمایه‌گذاری برای سرمایه‌گذار جذابتر است زیرا میزان بیشتری ارز می‌تواند از آن کسب کند و اگر این میزان کم باشد دیگر کسی مایل به سرمایه‌گذاری در این صندوق نیست زیرا دیگر آورده‌ای برای اون ندارد.

**هدف** از این پروژه پیاده‌سازی سیستم رمزارزهای اعتبارات کربن با استفاده از یادگیری ماشینی به عنوان عامل تبدیل توکن‌های غیر قابل تعویض به قابل تعویض می‌باشد

## 5-1 ساختار پروژه

در فصل‌هایی که در ادامه می‌آیند در فصل دوم به بیان سیستم‌های فعلی برای بازار رمزارزهای اعتبار کربن پرداخته و سپس در فصل‌های سوم به راهکارهای یادگیری تقویتی پرداخته و به نقاط ضعف و قدرت هر یک از آن‌ها برای پروژه خود می‌پردازیم. در فصل چهارم نیز راهکار خود را ارائه نموده و نتیجه‌گیری می‌کنیم.



## فصل دوم: سیستم‌های رمزارزی اعتبار کرین موجود

## 1-2 مقدمه

در حوزه سیستم‌های رمزارزهای اعتبارات کربن دو محصول تجاری به حد مطلوب رسیده‌اند. یکی سیستم توکان<sup>5</sup> و دیگری سیستم انرژی وب اریجین<sup>6</sup> در دو بخش به هر یک از آن‌ها و مزایا و معایب آن‌ها می‌پردازیم.

## 2-2 بررسی سیستم توکان

### 1-2-2 بازار انرژی توکان

توکان به ارائه زیرساخت‌های دیجیتال برای اعتبارهای کربنی توکنیزه شده می‌پردازد. هدف اصلی این سایت، توسعه داوطلبانه کربن<sup>6</sup> با شفافیت و اعتبار بالا است. در این راستا، از فناوری بلاکچین برای تضمین شفافیت و صداقت در معاملات اعتبار کربنی استفاده می‌شود.

تمام اطلاعات مربوط به اعتبار کربنی در یک پایگاه داده مبتنی بر بلاکچین ثبت می‌شوند که این امر به همگان اجازه می‌دهد تا به صورت مستقل از ویژگی‌ها، تاریخچه معاملات و داده‌های قیمتی هر اعتباری مطلع شوند. توکان ارتباط نزدیکی با متخصصان، سازمان‌های ثبتی و دیگر نهادهای صنعتی دارد تا از مزایای فناوری‌های جدید بهره‌مند شوند، در حالی که به دقت خطرات را مدیریت می‌کنند.

زیرساخت‌های توکن تحت قوانین سوئیس فعالیت می‌کنند و برنامه‌هایی برای اجرای تدابیر احراز هویت مشتری<sup>7</sup> به منظور حفظ یکپارچگی اعتبارهای کربنی توکنیزه شده دارند. این زیرساخت‌ها اعتبار کربنی را با برنامه‌های کاربردی بلاکچین متصل می‌کنند که نتیجه آن امکانات نامحدود برای نوآوری در نسل بعدی محصولات و انگیزه‌های مثبت برای اقلیم است تا اقدامات آب و هوایی را در مقیاس وسیعی فعال کنند.

از زمان راه اندازی توکان در اکتبر 2021، اتفاقات بزرگی در زمینه بازنشستگی اعتبار کربنی رخ داده است. پلتفرم برای کاربران متنوعی طراحی شده است، از جمله کسانی که به دنبال نظارت بر فعالیت بازار یا استفاده از داده‌های معاملاتی قابل دسترسی عمومی هستند.

---

<sup>6</sup> Voluntary Carbon Market (VCM)

<sup>7</sup> Know Your Customer (KYC)

## فصل دوم: مفاهیم پایه و کارهای مرتبط

در بخش توصیفات، سازمان‌های مختلفی به تایید نقش توکان در مقابله با تغییرات آب و هوایی با استفاده از فناوری وب ۳ اشاره کرده‌اند. محصولات اصلی سایت شامل پل‌های کربنی برای توکنیزه<sup>8</sup> کردن اعتبارهای کربنی موجود در ثبت‌های کربنی معتبر، استخرهای کربنی و بازنشستگی کربنی است که امکان خرید و بازنشستگی اعتبار کربنی را با شفافیت فراهم می‌کند.

در نهایت، مأموریت اصلی توکان ایجاد زیرساخت‌های مالی برای هدایت تریلین‌ها دلار به سمت بهترین راحل‌های آب و هوایی است. سایت توکان همچنین از همکاری‌های خود با سازمان‌ها و نهادهای مختلف در حوزه اقدامات آب و هوایی و فضای بلاکچین یاد می‌کند.

پل کربن توکان یک ابتکار نوآورانه است که به افراد و سازمان‌ها امکان می‌دهد تا اعتبارات کربنی خود را به شکل توکن‌های دیجیتال در بلاکچین قرار دهند. این فرآیند تبدیل، اعتبارات کربنی سنتی را به دارایی‌های دیجیتالی تبدیل می‌کند که می‌توانند به راحتی در بازارهای مالی دیجیتال مورد استفاده و معامله قرار گیرند. این تحول، فرصت‌های جدیدی را برای سرمایه‌گذاری و مدیریت اعتبارات کربنی فراهم می‌کند.

یکی از مهمترین ویژگی‌های توکن‌های کربنی، شفافیت آن‌ها است. با استفاده از فناوری بلاکچین<sup>9</sup> هر تراکنش و نگهداری این توکن‌ها به طور کامل قابل ردیابی و شفاف است. این ویژگی به افزایش اعتماد و شفافیت در بازار کربن کمک می‌کند و اطمینان می‌دهد که اعتبارات کربنی به درستی مورد استفاده قرار می‌گیرند. علاوه بر شفافیت، توکن‌سازی اعتبارات کربنی امکان برنامه‌ریزی و تنظیم دقیق این دارایی‌ها را فراهم می‌کند. این امر به سازمان‌ها و فعالان محیط زیست اجازه می‌دهد تا استراتژی‌های پیچیده‌تری را برای مدیریت اعتبارات کربنی خود طراحی کنند. همچنین، توکن‌های کربنی می‌توانند به راحتی تقسیم شوند<sup>9</sup> که این امر دسترسی به بازار کربن را برای سرمایه‌گذاران کوچکتر و فردی ساده‌تر می‌کند.

یکی دیگر از مزایای مهم توکن‌سازی کربن «ترکیب‌پذیری آن با اکوسیستم مالی غیرمتمرکز<sup>9</sup> است. این ویژگی به توکن‌های کربنی اجازه می‌دهد تا در محصولات و خدمات مالی مختلف مورد استفاده قرار گیرند» از جمله در پروژه‌های سرمایه‌گذاری پایدار و تأمین مالی سبز.

در نهایت، پل کربن توکان بخشی از تلاش‌های گسترده‌تر برای آوردن دارایی‌های محیطی به فضای دیجیتال است. این ابتکار عمل به دنبال تسهیل دسترسی، مدیریت و استفاده از اعتبارات کربنی در مبارزه با تغییرات آب و هوایی است و امکانات جدیدی را برای سرمایه‌گذاری مسئولانه و پایدار در زمینه محیط زیست فراهم می‌کند.

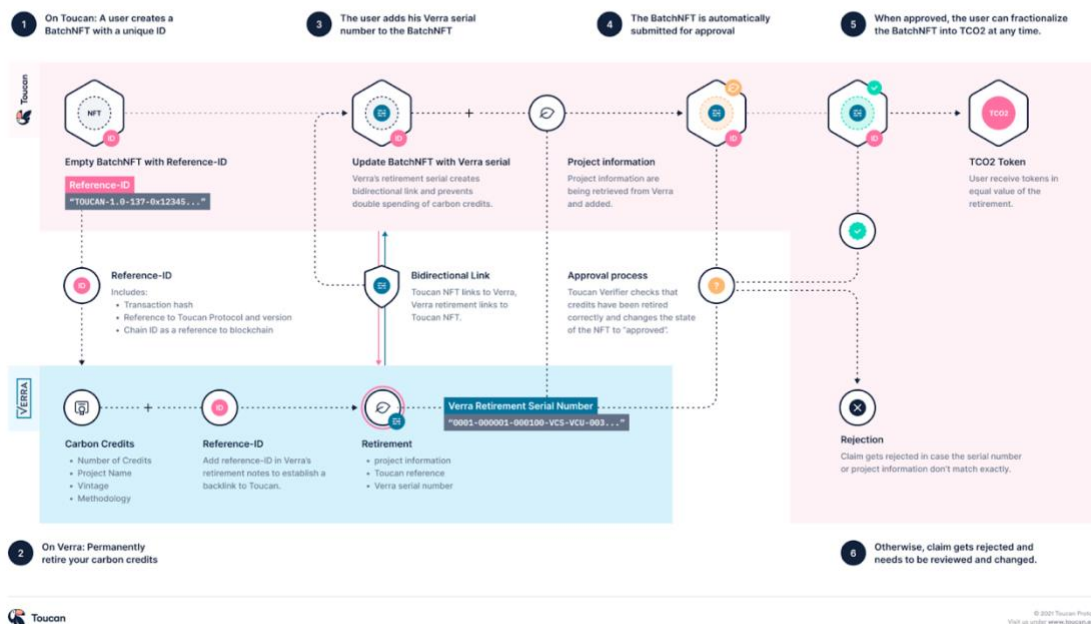
---

<sup>8</sup> Tokenization

<sup>9</sup> Decentralized Finance (Defi)

## 2-2-2 سیستم پل در توکان

### How does the Carbon Bridge work?



تصویر 2 شرح پل توکان

تصویر نشان‌دهنده نحوه کار "پل کربن" است. مراحل مختلف این فرآیند به شرح زیر است:

1. در توکان: یک کاربر یک توکن غیر قابل تعویض (دسته‌ای با یک شناسه‌ی منحصر به فرد ایجاد می‌کند. این توکن خالی با یک "شناسه‌ی مرجع" ارائه می‌شود. شناسه‌ی مرجع شامل اطلاعاتی مانند هش معامله، ارجاع به پروتکل توکان و نسخه و همچنین شناسه‌ی زنجیره‌ی بلوکی است.
2. در ورا<sup>10</sup> می‌توان اعتبارات کربنی خود را برای همیشه بازنشسته کرد. این شامل تعداد اعتبارات، نام پروژه، و روش است.
3. کاربر شماره‌ی سریال ورا خود را به توکن دسته‌ای اضافه می‌کند. وقتی شماره‌ی سریال ورا به توکن اضافه می‌شود. یک پیوند دوسویه ایجاد می‌شود که جلوی مصرف دو برابر اعتبارات کربن را می‌گیرد.

<sup>10</sup> Verra

## فصل دوم: مفاهیم پایه و کارهای مرتبط

۴. توکن غیر قابل عوض دسته‌ای به طور خودکار برای تأیید ارسال می‌شود. اطلاعات پروژه از ورا بازیابی شده و اضافه می‌شود. فرآیند تأیید شروع می‌شود. در اینجا سازمان توکان اطمینان حاصل می‌کند که اعتبارات به درستی بازنشسته شده‌اند و وضعیت توکن غیر قابل عوض به "تایید شده" تغییر می‌دهد.

۵. وقتی تأیید شده است، کاربر می‌تواند توکن غیر قابل عوض دسته‌ای را در هر زمان به توکن‌های قابل عوض با امکان نقدینگی کم با نام TCO2 تجزیه کند. کاربران توکن‌ها را با ارزش معادل بازنشستگی دریافت می‌کنند.

6. در غیر این صورت، اگر اطلاعات دقیقاً مطابقت نداشته باشد، ادعا رد می‌شود و باید مورد بازبینی و تغییر قرار گیرد.

برخی نکات کلیدی عبارتند از:

- توکنیزاسیون اعتبارات کربن: پس از توکنیزه و تجزیه شدن اعتبارات کربن یک پروژه با استفاده از پل کربن» این

اعتبارات به صورت توکن‌های TCO2 نمایش داده می‌شوند.

- ویژگی‌های مخصوص پروژه: هر قرارداد توکن TCO2 تمام ویژگی‌های مخصوص پروژه را حفظ می‌کند x

مانند کشور یا روش‌شناسی پروژه.

- عدم قابلیت تبادل: توکن‌ها از یک پروژه با توکن‌ها از پروژه دیگر قابل تبادل نیستند. حتی اگر بسیار شبیه به هم باشند.

هدف از این رویکرد، ساده‌سازی فرآیند معامله با اعتبارات کربن و ایجاد یک بازار استاندارد و نقدی‌تر برای آن‌ها است. ویژگی‌های یک توکن کربن می‌تواند به صورت زیر باشد:

standard = Puro — only accepts TCO2s from the Puro standard

vintage = >2021 — only accepts TCO2s from 2021 and later.

country = Colombia — only accepts TCO2s from Colombia.

## 3-2-2 بررسی استخرها در سیستم توکان

استخرهای کربن در واقع روشی برای گروه‌بندی چندین تن کربن توکنیزه شده مخصوص پروژه (توکن‌های  $\text{TCO}_2$ ) در قالب توکن‌های استخر قابل تبادل هستند. این سیستم به منظور تسهیل در نقدینگی و کشف قیمت برای کلاس‌های مختلف دارایی‌های کربنی طراحی شده است.

استخرهای کربن، چندین تن کربن دی‌اکسید ( $\text{TCO}_2$ ) توکنیزه شده خاص پروژه را به توکن‌های استخر قابل تبدیل ترکیب می‌کنند. این امر نقدینگی و کشف قیمت را برای انواع مختلف دارایی‌های کربنی فراهم می‌کند.

### نحوه عملکرد استخرهای کربن

پس از اینکه اعتبارات کربن یک پروژه با استفاده از پل کربن توکنیزه و خرد شد، این اعتبارات به عنوان  $\text{TCO}_2$  توکن‌ها (تن‌های کربن توکنیزه شده) نشان داده می‌شوند. قرارداد یک توکن  $\text{TCO}_2$  همچنان حاوی تمام ویژگی‌های خاص پروژه، مانند کشور یا روش‌شناسی پروژه است. این بدان معناست که توکن‌های یک پروژه با توکن‌های پروژه دیگر قابل تعویض نیستند، حتی اگر بسیار مشابه باشند.

سپس توکن‌های  $\text{TCO}_2$  می‌توانند در استخرها سپرده شوند که نقدینگی و قابلیت تبدیل را در سراسر انواع پروژه‌های مشابه بهبود می‌بخشند.

### ترجمه فارسی

استخرهای کربن یک فناوری جدید هستند که برای تسهیل تجارت دارایی‌های کربن طراحی شده‌اند. آنها با ترکیب چندین تن کربن دی‌اکسید ( $\text{TCO}_2$ ) توکنیزه شده خاص پروژه به توکن‌های استخر قابل تبدیل، نقدینگی و کشف قیمت را برای انواع مختلف دارایی‌های کربنی فراهم می‌کنند.

عملکرد استخرهای کربن به شرح زیر است:

ابتدا، اعتبارات کربن یک پروژه با استفاده از یک پلتفرم توکن‌سازی مانند Carbon Bridge توکنیزه می‌شوند. این بدان معناست که آنها به عنوان یک دارایی دیجیتالی منحصر به فرد که می‌تواند در بازارهای مالی معامله شود، نمایندگی می‌شوند.

سپس، این اعتبارات کربن توکنیزه شده می‌توانند در یک استخر کربن سپرده شوند. استخر کربن یک قرارداد هوشمند است که مدیریت توکن‌های کربن را در یک گروه خاص کنترل می‌کند.

## فصل دوم: مفاهیم پایه و کارهای مرتبط

با سپرده‌گذاری توکن‌های کربن در استخر، آنها به توکن‌های استخر قابل تبدیل تبدیل می‌شوند. توکن‌های استخر قابل تبدیل از نظر ارزش برابر هستند و می‌توانند با یکدیگر معامله شوند.

استخرهای کربن می‌توانند مزایای متعددی برای بازارهای کربن داشته باشند. آنها می‌توانند:

- نقدینگی بازار کربن را افزایش دهند.
- کشف قیمت را بهبود بخشند.
- سرمایه‌گذاری در پروژه‌های کاهش انتشار کربن را آسان‌تر کنند.

استخرهای کربن یک فناوری جدید و نوظهور هستند که پتانسیل تغییر بازارهای کربن را دارند. آنها می‌توانند به کاهش انتشار کربن و مبارزه با تغییرات آب و هوایی کمک کنند.

## 3-2 بررسی سیستم انرژی وب اریجین

انرژی وب اریجین یک شبکه بلاکچینی مخصوص صنعت انرژی است که به منظور ارتقاء انتقال و مدیریت انرژی تجدیدپذیر و پایدار توسعه یافته است. این پروژه به عنوان یک شبکه برای اشیاء 11 انرژی بازارهای جدید را فراهم می‌کند و از تکنولوژی بلاکچین برای تضمین امنیت، شفافیت، و کارایی در انتقال اطلاعات و انرژی استفاده می‌کند.

مهمترین اهداف انرژی وب اریجین عبارتند از:

1. ایجاد بازار انرژی دیجیتال: انرژی وب اریجین به تشکیل بازارهای انرژی دیجیتال کمک می‌کند که انرژی تولیدی از منابع تجدیدپذیر مانند باد، خورشید و سایر منابع پایدار را به تولیدکنندگان و مصرف‌کنندگان انرژی متصل کند.
2. افزایش شفافیت: با استفاده از تکنولوژی بلاکچین، تمامی تراکنش‌ها و معاملات مربوط به انتقال انرژی به شفافیت بیشتری دست خواهند یافت و اطلاعات دقیقی از میزان تولید و مصرف انرژی در دسترس قرار خواهد گرفت.
3. امنیت: امنیت اطلاعات و انتقال انرژی در انرژی وب اریجین با استفاده از تکنولوژی بلاکچین تقویت می‌شود، که از حملات سایبری و تغییرات غیرمجاز جلوگیری می‌کند.

---

<sup>11</sup> Internet of Things (IOT)

## فصل دوم: مفاهیم پایه و کارهای مرتبط

4. ایجاد سیستم‌های انرژی هوش مصنوعی: انرژی وب اریجین به توسعه سیستم‌های هوش مصنوعی که

بهینه‌سازی مصرف انرژی و مدیریت آن را بهبود می‌بخشند، کمک می‌کند.

5. حمایت از تجدیدپذیری و پایداری: این پروژه از تولید انرژی از منابع تجدیدپذیر و کاهش انتشار

گازهای گلخانه‌ای حمایت می‌کند.

انرژی وب اریجین یکی از پروژه‌های مهم در زمینه بلاکچین در صنعت انرژی است که به منظور بهبود

مدیریت و انتقال انرژی تجدیدپذیر و پایدار به کار می‌رود.

انرژی وب اریجین یک مجموعه از کیت‌های توسعه نرم‌افزاری<sup>12</sup> و سرویس‌های سمت سرویس‌دهنده<sup>13</sup> است

که در مجموع یک پلتفرم برای صدور، مدیریت و معامله گواهینامه‌های ویژگی انرژی<sup>14</sup> ارائه می‌کنند.

گواهینامه ویژگی انرژی یک سند رسمی است که تضمین می‌کند که انرژی تولید شده از منابع تجدیدپذیر می‌آید.

استانداردهای مختلفی وجود دارند که نحوه ذخیره و اعتبارسنجی داده‌ها را تنظیم می‌کنند. در اروپا، این سند به نام

«گواهی اصالت»<sup>15</sup> شناخته می‌شود، در شمال آمریکا «گواهینامه انرژی تجدیدپذیر»<sup>16</sup> نام دارد و در بخش‌هایی

از آسیا، آفریقا، خاورمیانه و آمریکای لاتین استاندارد حاکم بر آن «گواهی انرژی تجدیدپذیر بین‌المللی»<sup>17</sup> است.

استانداردها ممکن است متفاوت باشند، اما همه از اصول اساسی مشترکی بهره می‌برند.

قصد اصلی گواهینامه‌های ویژگی انرژی اثبات آن است که انرژی مصرف شده از منابع تجدیدپذیر به دست آمده

است. گواهینامه‌های ویژگی انرژی می‌توانند در یک بازار معاملاتی معامله و خریداری شوند.

### 2-3-1 بررسی معماری انرژی وب اریجین

---

<sup>12</sup> Software Development Kit (SDK)

<sup>13</sup> Backend

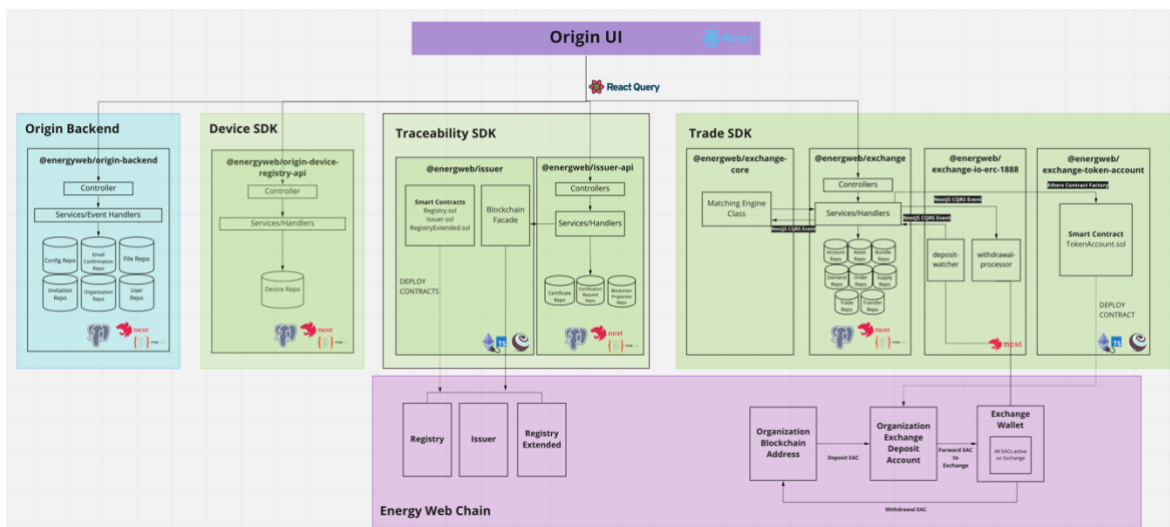
<sup>14</sup> Energy Attribute Certificate (EAC)

<sup>15</sup> Guarantee of Origin (GO)

<sup>16</sup> Renewable Energy Certificate (REC)

<sup>17</sup> International-Renewable Energy Certificate (I-REC)





تصویر 3 - معماری انرژی وب اریجین

این تصویر یک معماری نرم‌افزاری را نشان می‌دهد که برای پلتفرمی با نام "Energy Web Chain" طراحی شده است. پلتفرم به چند بخش اصلی تقسیم شده است که هر کدام دارای ماژول‌ها و کلاس‌های مختلفی هستند. در اینجا به شرح هر بخش می‌پردازیم:

1. **پشتگاه<sup>18</sup> اریجین:** این بخش شامل یک کنترلر و سرویس‌ها/رسیدگی‌کننده‌های رویداد است که به پیکربندی، ایمیل، فایل، دعوت‌نامه، سازمان و مخزن کاربران مرتبط می‌شود. این قسمت از نرم‌افزار عمدتاً برای مدیریت پشتیبانی و پیکربندی سیستم استفاده می‌شود.
2. **وسیله کیت توسعه نرم‌افزار:** این بخش نیز شامل یک کنترلر و سرویس‌هایی است که به یک مخزن دستگاه متصل هستند. این کیت توسعه نرم‌افزار ممکن است برای تعامل با دستگاه‌های فیزیکی در شبکه و ثبت اطلاعاتشان در سیستم استفاده شود.
3. **کیت توسعه نرم‌افزار ردیابی:** این بخش شامل کنترلرها و سرویس‌هایی است که با قراردادهای هوشمند و یک نمای بلاکچین ارتباط دارند. قراردادهای هوشمند شامل ثبت‌نام‌ها و صدور گواهی‌نامه‌ها هستند. این بخش برای ردیابی و اعتبارسنجی تراکنش‌ها و اسناد در شبکه استفاده می‌شود.

<sup>18</sup> Backend

## فصل دوم: مفاهیم پایه و کارهای مرتبط

4. کیت توسعه نرم افزار معامله: این بخش شامل یک موتور مطابقت و کنترلرها و سرویس ها/رسیدگی کننده هایی است که به حساب ها، دارایی ها، سفارش ها و معاملات مرتبط هستند. این بخش به نظر می رسد که برای تسهیل معاملات و تبادل های تجاری در پلتفرم طراحی شده است.

شامل سه بخش اصلی است: ثبت<sup>19</sup>، صادرکننده و ثبت گسترده<sup>20</sup> است. این Energy Web Chain قسمت ها ممکن است از بلاکچین هستند که برای ثبت و صدور گواهینامه ها و سایر اسناد مورد استفاده قرار می گیرند.

### 1-1-2-3 کیت توسعه نرم افزار ردیابی

کیت توسعه نرم افزار ردیابی ردیابی کیت توسعه نرم افزار مسئول درخواست، صدور و مبادله گواهینامه های ویژگی انرژی در پلتفرم اریجین است. گواهینامه ها توسط صاحبان دستگاه درخواست می شوند و توسط صادرکنندگان تأیید می شوند. صادرکننده نهادی است که مسئول بررسی شواهد تولید انرژی و صدور گواهینامه به گونه ای است که با استانداردهای قانونی و صنعتی مطابقت داشته باشد.

کیت توسعه نرم افزار ردیابی دارای دو بسته اصلی است:

صادرکننده

بسته صادرکننده مدیریت گواهینامه ها را در بلاک چین انجام می شود و دارای دو مؤلفه اصلی است:

- قراردادهای هوشمند که چرخه عمر گواهینامه ها را در زنجیره بلوک حفظ می کنند.
- نمایه ها برای تعامل با توابع قرارداد هوشمند. صادرکننده از این نمایه ها استفاده می کند تا به جای تعامل مستقیم با بلاک چین، از آنها استفاده کند.

API صادرکننده:

گواهینامه های ویژگی انرژی توسط نهادهای رسمی به دستگاه های تولیدکننده برق صادر می شوند تا گواهی کنند که در بازه زمانی مشخص شده، مقدار مشخصی انرژی سبز تولید کرده اند. نهادهای صادرکننده رسمی برای مناطق مختلف وجود دارند. گواهینامه ها باید اطلاعات دستگاه تولید، حجم کل انرژی تولید شده و بازه زمانی که انرژی در آن تولید شده را در نظر بگیرند. حجم انرژی تولید شده سپس می تواند به واحدهای کوچکتر تقسیم شود و به عنوان گواهینامه های ویژگی انرژی به خریداران مختلف فروخته شود. توجه داشته باشید که رابط پیاده سازی مرجع

<sup>19</sup> Registry

<sup>20</sup> Registry Extended

## فصل دوم: مفاهیم پایه و کارهای مرتبط

Origin از Mwh به عنوان واحد استاندارد حجم انرژی استفاده می‌کند، اما هر واحدی می‌تواند بسته به نیازهای پیاده‌سازی استفاده شود. گواهینامه‌هایی که صادر می‌شوند:

- منحصر به فرد هستند.
- قابل ردیابی هستند.
- قابل تأیید هستند.

ردیابی SDK یک ابزار قدرتمند برای مدیریت گواهینامه‌های ویژگی انرژی در بلاک چین است. این SDK اطمینان حاصل می‌کند که گواهینامه‌ها منحصر به فرد، قابل ردیابی و قابل تأیید هستند.

### 2-3-1-2 بررسی کیت توسعه نرم‌افزار معامله

کیت توسعه نرم‌افزار معامله برای سازگاری با کیت توسعه نرم‌افزار ردیابی طراحی شده است که یک کیت توسعه نرم‌افزار دیگر برای پلتفرم اریجین است. کیت توسعه نرم‌افزار ردیابی مسئول مدیریت چرخه عمر گواهینامه‌های ویژگی انرژی در بلاک‌چین است. با ادغام با کیت توسعه نرم‌افزار ردیابی، کیت توسعه نرم‌افزار تجارت می‌تواند اطمینان حاصل کند که تنها گواهینامه‌های ویژگی انرژی معتبر در صرافی معامله می‌شوند.

کیت توسعه نرم‌افزار تجارت دارای چهار بسته اصلی است:

هسته صرافی: این بسته عملکرد اصلی صرافی را ارائه می‌دهد، مانند مطابقت سفارشات و تسویه حساب.

صرافی: این بسته رابط سطح بالاتری را برای ساخت صرافی‌ها ارائه می‌دهد. این پیچیدگی بسته هسته صرافی را خلاصه می‌کند و توسعه صرافی‌های سفارشی را آسان‌تر می‌کند.

حساب توکن صرافی: این بسته راهی برای مدیریت گواهینامه‌های ویژگی انرژی در صرافی ارائه می‌دهد. این به توسعه‌دهندگان اجازه می‌دهد تا تعادل گواهینامه ویژگی انرژی را ایجاد، به‌روزرسانی و حذف کنند.

Exchange IO ERC1888: این بسته راهی برای تعامل با استاندارد ERC1888 برای مبادلات اتمی ارائه می‌دهد. مبادلات اتمی نوعی تراکنش است که به دو طرف اجازه می‌دهد تا توکن‌ها را در بلاک‌چین‌های مختلف بدون نیاز به واسطه مورد اعتماد مبادله کنند.

کیت توسعه نرم‌افزار تجارت یک ابزار قدرتمند است که می‌توان از آن برای ساخت انواع صرافی‌های سفارشی برای مدیریت گواهینامه‌های ویژگی انرژی استفاده کرد. این برای استفاده آسان و ادغام با سایر کیت‌های توسعه نرم‌افزار طراحی شده است.

مزایای استفاده از کیت توسعه نرم‌افزار تجارت:

## فصل دوم: مفاهیم پایه و کارهای مرتبط

- کیت توسعه نرم افزار تجارت یک ابزار قدرتمند و انعطاف پذیر است که می توان از آن برای ساخت انواع صرافی های سفارشی برای گواهینامه های ویژگی انرژی استفاده کرد.
  - کیت توسعه نرم افزار تجارت با کیت توسعه نرم افزار ردیابی برای اطمینان از معامله تنها گواهینامه های ویژگی انرژی معتبر ادغام شده است.
  - کیت توسعه نرم افزار تجارت برای سازگاری با سایر سیستم های مبتنی بر بلاکچین طراحی شده است.
- کاربردهای کیت توسعه نرم افزار تجارت:
- کیت توسعه نرم افزار تجارت می تواند برای طیف وسیعی از کاربردها استفاده شود، از جمله:
  - تسهیل مبادلات گواهینامه ویژگی انرژی بین تولیدکنندگان برق تجدیدپذیر و خریداران
  - ایجاد بازارهای ثانویه برای گواهینامه های ویژگی انرژی
  - ردیابی و گزارش تولید برق تجدیدپذیر
- کیت توسعه نرم افزار تجارت یک ابزار ارزشمند برای توسعه دهندگانی است که می خواهند در زمینه انرژی تجدیدپذیر مشارکت کنند.

## 4-2 جمع بندی

در کل می توان مزایا و معایب دو سیستم را به صورت جدول زیر خلاصه کرد.

جدول 1- مقایسه توکان و انرژی وب اریجین

توکان	انرژی وب اریجین
محوریت پل کربن برای تبدیل به TC02.	تنها اعتبار توکنیزه می شود.
محوریتی برای انتساب اعتبار به دستگاه ندارد.	اعتبار کربن را به دستگاه تولید انرژی سبز منتسب می کند.

## فصل دوم: مفاهیم پایه و کارهای مرتبط

محدود به تولید انرژی سبز نیست. بلکه مقابله با جنگل‌زدایی را نیز در دسته اعتبارات خود می‌داند.	اعتبار کربن را محدود به تولید انرژی سبز کرده.
امکان تولید رمزارز بنا به میزان تولید انرژی سبز به صورت آنلاین را ندارد.	بالقوه می‌تواند رمزارز را با توجه به میزان تولید انرژی سبز در دستگاه‌های تولید با کمک IOT و فناوری Oracle تولید کند.
پروژه کماکان با توجه به دشواری‌های سر راهش فعال است.	توسعه پروژه متوقف شده و آخرین نسخه آن نیز قابل مستقرسازی <sup>21</sup> و اجرا نیست.

ساختار پروژه توکان برای پروژه ما مطلوب‌تر بوده با این تفاوت که پروژه ما به جای استخرهای پروژه توکان که توکن‌های TCO2 را یکبار فقط به توکن‌های با نقدینگی بیشتر تبدیل کند، آن‌ها را در خود نگه داشته و با توجه به شرایط متفاوت با فرکانس‌های متفاوت توکن‌های با به نقدینگی بالا تولید می‌کند. یعنی استخرها در این پروژ معادل صندوق سرمایه‌گذاری رمزارزی خواهند بود. خال مسئله اصلی فرکانس تولید توکن با نقدینگی بالاست. این فرکانس باید تعادل بین تمایل مردم به سرمایه‌گذاری بیشتر و نقدینگی و قیمت توکن را حفظ کند.

---

<sup>21</sup> Deployment

## فصل سوم: بررسی روش‌های یادگیری تقویتی

### 3-1 مقدمه

یادگیری تقویتی آن است که چه کاری انجام دهیم؛ چگونه موقعیت‌ها را به اقدامات نگاشت کنیم؛ تا حداکثر سیگنال پاداش عددی را بدست آورده شود. مانند بسیاری از اشکال یادگیری ماشینی، به یادگیرنده گفته نمی‌شود که کدام اقدامات را انجام دهد، بلکه باید با امتحان آنها کشف کند که کدام اقدامات بیشترین پاداش را می‌دهند. در مواردی که جالب و چالش برانگیز هستند، اقدامات ممکن است نه تنها پاداش فوری، بلکه موقعیت بعدی و از طریق آن تمام پاداش‌های بعدی را نیز تحت تأثیر قرار دهند. این دو ویژگی - جستجو و خطا و پاداش تأخیری - دو ویژگی مهم متمایز یادگیری تقویتی هستند.

یادگیری تقویتی با مشخص کردن الگوریتم‌های یادگیری تعریف نمی‌شود، بلکه با مشخص کردن یک مسئله یادگیری تعریف می‌شود. هر الگوریتمی که برای حل آن مشکل مناسب باشد را یک الگوریتم یادگیری تقویتی می‌دانیم. ایده اصلی این است که مهمترین جنبه‌های مشکل واقعی را با استفاده از تعامل با محیط خود که خود یک عامل یادگیری است، برای دستیابی به هدف مشخصی به کار گیریم. چنین عاملی باید تا حدی بتواند وضعیت محیط را حس کند و باید بتواند اقداماتی انجام دهد که بر آن وضعیت تأثیر بگذارد. عامل همچنین باید یک هدف یا اهدافی مربوط به وضعیت محیط داشته باشد. [7]

روش‌های مختلفی از یادگیری ماشینی ارائه شده که بیان هر یک از آنها می‌پردازیم.

### 3-2 بررسی روش‌های مختلف یادگیری تقویتی

انواع مختلف از الگوریتم‌ها و رویکردها در تقویت یادگیری وجود دارد، از جمله:

#### 3-2-1 تقویت یادگیری بدون مدل

یادگیری تقویتی بدون مدل<sup>22</sup> یک روش در حوزه یادگیری ماشین است که برای حل مسائل تصمیم‌گیری در محیط‌های تعاملی استفاده می‌شود. در این نوع از یادگیری تقویتی، ما به جای استفاده از یک مدل دقیق از محیط برای پیش‌بینی آینده، تلاش می‌کنیم تا از تجربیات مستقیم با محیط استفاده کنیم تا یک استراتژی بهینه برای تصمیم‌گیری در محیط توسعه دهیم.

---

<sup>22</sup> Model Free Learning

در یادگیری تقویتی بدون مدل، یک عامل در یک محیط قرار دارد و در هر مرحله از زمان، عامل اقدامی انجام می‌دهد و از محیط پاداشی دریافت می‌کند. هدف این عامل این است که استراتژی بهینه‌ای را کشف کند که به ماکسیم کردن جمع ارزش (مجموع پاداش‌ها در طول زمان) منجر می‌شود.

روش‌های یادگیری تقویتی بدون مدل شامل الگوریتم‌هایی مانند Q-Learning ، Sarsa و AC3 می‌شوند. این الگوریتم‌ها بر اساس تجربیات جمع‌آوری شده در محیط، توابعی مانند تابع ارزش یا تابع عمل را برای انتخاب عمل بهینه یا تخمین ارزش وضعیت‌ها و عمل‌ها آموزش می‌دهند.

مزیت اصلی یادگیری تقویتی بدون مدل این است که به عامل اجازه می‌دهد تا در محیط‌های پیچیده و غیرقطعی عمل کند و به صورت تجربی به یادگیری بپردازد، بدون نیاز به دانستن مدل دقیقی از محیط. این الگوریتم‌ها معمولاً با استفاده از تکنیک‌هایی مانند تابع‌های مشتق‌گیری و نمونه‌برداری از تجربیات، به عامل اجازه می‌دهند تا به تدریج استراتژی بهینه‌ای را کشف کند و بهبودش دهد.

### 3-2-1-1 یادگیری Q

یادگیری Q<sup>23</sup> یکی از الگوریتم‌های مهم در حوزه یادگیری تقویتی است که برای یادگیری یک استراتژی بهینه برای تصمیم‌گیری در محیط‌های تعاملی استفاده می‌شود. این الگوریتم به ویژه برای مسائلی که محیط آن‌ها تقریباً قطعی<sup>24</sup> استفاده می‌شود، اما می‌توان آن را به محیط‌های غیرقطعی<sup>25</sup> نیز تعمیم داد.

در یادگیری Q ، عامل یک جدول به نام "Q-Table" را تعریف می‌کند. این جدول دارای ابعادی معادل تعداد وضعیت‌ها و تعداد عمل‌ها در محیط است. هر خانه از این جدول مقداری به نام Q-value دارد که نشان‌دهنده ارزش ترکیبی از وضعیت و عمل است.

عامل به صورت تجربی و در طول زمان اقداماتی در محیط انجام می‌دهد و پاداش‌های دریافتی را مشاهده می‌کند. با استفاده از این تجربیات، عامل مقادیر Q-value را به روز می‌کند تا بهبود استراتژی خود برای انتخاب عمل‌ها را دست‌یابی کند.

$$Q(s, a) = Q(s, a) + \alpha \times [R(s, a) + \gamma \times \max_{a'} Q(s', a') - Q(s, a)]$$

<sup>23</sup> Q Learning

<sup>24</sup> Deterministic

<sup>25</sup> Stochastic



در این فرمول:

- $Q(s, a)$  : برای وضعیت  $s$  و عمل  $a$ .
- $\alpha$  (نرخ یادگیری) : یک پارامتر که نشان‌دهنده میزان اهمیت تجربیات جدید است و می‌تواند بین 0 و 1 باشد.
- $R(s, a)$  : پاداش دریافتی برای انجام عمل  $a$  در وضعیت  $s$ .
- $\gamma$  (عامل تخفیف) : یک پارامتر که نشان‌دهنده تاثیر پاداش‌های آینده است و می‌تواند بین 0 و 1 باشد.
- $\max(Q(s', a'))$  : بیشترین Q-value برای وضعیت بعدی  $s'$  و همه عمل‌های ممکن در آن وضعیت.
- $Q(s, a)$  : مقدار Q-value قبلی برای وضعیت  $s$  و عمل  $a$ .

عامل با استفاده از این فرمول مقدار Q-value را به روزرسانی کرده و به تدریج استراتژی بهینه‌ای برای تصمیم‌گیری در محیط را یاد می‌گیرد. این فرآیند ادامه پیدا می‌کند تا مقادیر Q-value به تقریب به مقادیر بهینه برای استراتژی بهینه نزدیک شوند. یادگیری Q معمولاً به عنوان یک روش مدل‌نسبی برای یادگیری تقویتی بدون مدل شناخته می‌شود، زیرا بدون نیاز به دانش دقیق از مدل محیط، عامل می‌تواند به تدریج استراتژی بهینه‌ای را کشف کند. [8]

### SARSA - State-Action-Reward-State-Action 2-1-2-3

الگوریتم Sarsa یکی از الگوریتم‌های یادگیری تقویتی است که برای یادگیری یک استراتژی بهینه برای تصمیم‌گیری در محیط‌های تعاملی استفاده می‌شود. نام "Sarsa" از ترکیب اجتماعی اجزای اصلی این الگوریتم گرفته شده است:

S: State

وضعیت فعلی

A: Action

عمل انجام شده

R: Reward

پاداش دریافتی در ازای عمل

S': State

وضعیت بعدی

A': Action

عمل بعدی

الگوریتم Sarsa به صورت یک الگوریتم یادگیری Q-value عمل مدل می‌شود. عامل در هر مرحله از زمان، در وضعیت فعلی قرار دارد و یک عمل را انتخاب می‌کند. سپس عمل انتخاب شده را انجام می‌دهد، پاداشی را دریافت می‌کند و به وضعیت بعدی منتقل می‌شود. سپس عامل عمل بعدی را انتخاب می‌کند و فرآیند به این ترتیب ادامه می‌یابد.

روند آموزش الگوریتم Sarsa به این صورت است:

1. مقادیر اولیه برای Q-value ها  $(Q(s, a))$  مقداردهی می‌شود.
2. عامل در وضعیت فعلی (S) قرار دارد و یک عمل (A) را انتخاب می‌کند با استفاده از یک روش معین مانند  $\epsilon$ -greedy که امکان انتخاب تصادفی عمل را هم فراهم می‌کند.
3. عامل عمل انتخاب شده را انجام داده و پاداش (R) را دریافت می‌کند و به وضعیت بعدی (S') منتقل می‌شود.
4. سپس عامل در وضعیت بعدی (S') عمل بعدی (A') را بر اساس سیاست خود انتخاب می‌کند.
5. مقدار Q-value برای جفت وضعیت فعلی و عمل فعلی  $(Q(S, A))$  با استفاده از فرمول به روزرسانی زیر به‌روز می‌شود:

$$Q(s, a) = Q(s, a) + \alpha \times [R(s, a) + \gamma \times Q(s', a') - Q(s, a)]$$

که نمادها مانند الگوریتم یادگیری Q می‌باشد.

6. مراحل 2 تا 5 به صورت تکراری ادامه می‌یابد تا عامل یک استراتژی بهینه برای تصمیم‌گیری در محیط یاد بگیرد.

الگوریتم Sarsa به عنوان یک روش مدلی برای یادگیری تقویتی بدون مدل شناخته می‌شود، زیرا به تدریج تخمین بهینه Q-value ها را در محیط انجام می‌دهد و استراتژی بهینه را کشف می‌کند. [9]

### 3-1-2-3 روش‌های تقویت سیاست:

این روش‌ها به‌صورت مستقیم یک خط مشی که وضعیت‌ها را به اعمال نگاشت می‌کند، یاد می‌گیرند، به‌جای تخمین توابع ارزش. به عنوان مثال می‌توان به روش REINFORCE و روش‌های بازیگر-انتقادی (Actor-Critic) اشاره کرد.

از آن‌جا که این مدل یادگیری بنا به ذات گسسته خود با مسئله ما مطابقت ندارد بیش از این به آن نمی‌پردازیم.

### 3-2-2 یادگیری تقویتی با مدل

یادگیری تقویتی با مدل<sup>26</sup> یکی از روش‌های یادگیری تقویتی است که در آن یک مدل داخلی از محیط به وسیله عامل ساخته می‌شود و از این مدل برای انجام تصمیم‌گیری‌ها و یادگیری سیاست بهینه استفاده می‌شود. این روش تفاوت‌های مهمی با روش‌های دیگر یادگیری تقویتی دارد. مانند یادگیری تقویتی مبتنی بر خود در یادگیری تقویتی مبتنی بر مدل، مراحل کلی به شرح زیر هستند:

1. جمع‌داده: ابتدا، عامل با تعامل با محیط داده‌های تجربی جمع‌آوری می‌کند. این داده‌ها شامل اطلاعاتی از تجربیات عامل در محیط و نتایج اعمال اقدامات مختلف می‌شوند.

2. ساخت مدل: در مرحله دوم، عامل یک مدل داخلی از محیط ایجاد می‌کند. این مدل معمولاً یک مدل پیش‌بینی<sup>27</sup> نامیده می‌شود. این مدل به عامل اجازه می‌دهد تا پیش‌بینی‌هایی از نتایج اقدامات در محیط انجام دهد.

3. تولید سیاست: با استفاده از مدل داخلی محیط، عامل می‌تواند تعدادی تصمیم را به صورت آزمایشی تولید کند و ارزیابی کند که کدام یک از اقدامات بهینه‌ترین نتیجه را تولید می‌کند. این فرآیند معمولاً با استفاده از الگوریتم‌های بهینه‌سازی مانند الگوریتم‌های بهینه‌سازی تقریباً مطلق انجام می‌شود.

4. آموزش مدل: در مرحله آموزش مدل داخلی، عامل از داده‌های تجربی‌اش استفاده می‌کند تا مدل را بهبود دهد. این به معنای به روزرسانی پارامترهای مدل توسط الگوریتم‌های یادگیری مدل می‌باشد. این به عامل این امکان را می‌دهد که مدل داخلی خود را بهبود داده و تقریب بهتری از محیط ایجاد کند.

مزیت اصلی یادگیری تقویتی مبتنی بر مدل این است که به عامل امکان پیش‌بینی دقیق‌تری از تاثیر اقدامات خود در محیط می‌دهد و این اطلاعات به عامل کمک می‌کند تا به سرعت یادگیری کند. با این حال، این روش نیازمند

---

<sup>26</sup> Model-Based Reinforcement Learning

<sup>27</sup> Prediction Model

تولید و نگهداری یک مدل داخلی دقیق و معتبر از محیط است که ممکن است در بعضی موارد دشوار باشد. همچنین، نیاز به جمع‌آوری داده‌های تجربی از محیط نیز ممکن است هزینه و زمان بر باشد.

### 3-2-2-1 جستجوی درخت مونت کارلو<sup>28</sup>

یک الگوریتم تقویت یادگیری با مدل است که برای انجام تصمیمات در محیط‌هایی که عامل دارای میزان محدودی از اطلاعات است، استفاده می‌شود. این الگوریتم یک درخت از اقدامات ممکن و نتایج آنها را برای انتخاب تصمیمات ساخته و به‌روز می‌کند.

### 3-2-2-2 کنترل پیش‌بینی مدل<sup>29</sup>

از یک مدل یادگیری شده یا شبیه‌سازی شده از محیط برای پیش‌بینی وضعیت‌های آینده و بهینه‌سازی خطمشی کنترلی در یک بازه زمانی محدود استفاده می‌کند.

### 3-2-2-3 Dyna Q

همزمان از رویکردهای تقویت یادگیری بدون مدل و تقویت یادگیری با مدل استفاده می‌کند، با نگهداری از یک مدل از محیط برای شبیه‌سازی تجارب و به‌روزرسانی تابع ارزش.

یادگیری تقویتی با مدل معمولاً به مدل‌سازی محیط و تصمیم‌گیری در محیط‌های نامعین<sup>30</sup> توانمندی دارد. اگرچه لازم به ذکر است که این توانمندی باتوجه به شرایط محیط ممکن است متفاوت باشد. عموماً برای محیط‌های نامعین استفاده از این مدل از یادگیرنده‌ها نیازمند تلاش بیشتری است تا نتایج به حد مطلوب برسد.

### 3-2-3 یادگیری تقلیدی<sup>31</sup>

یادگیری تقویتی تقلیدی یک رویکرد در یادگیری ماشینی است که در آن عامل یادگیرنده سعی می‌کند عملکردی مشابه با عامل دیگری که به عنوان مرجع متخصص شناخته می‌شود، را تقلید کند. در این روش، عامل یادگیرنده نمونه‌هایی از عملکرد مرجع را مشاهده می‌کند و تلاش می‌کند تا این عملکرد را با تکرار تجربه و تمرین بهبود دهد.

مراحل یادگیری تقویتی تقلیدی عموماً به شرح زیر انجام می‌شود:

---

<sup>28</sup> Monte Carlo Tree Search - MCTS

<sup>29</sup> Model Predictive Control – MPC

<sup>30</sup> Stochastic

<sup>31</sup> Imitation Learning

1. جمع‌آوری داده‌های آموزشی: در این مرحله، داده‌هایی که توسط مرجع یا اکسپرت جمع‌آوری شده است به عامل یادگیرنده داده می‌شود. این داده‌ها معمولاً شامل توالی‌هایی از وضعیت‌ها و اقدامات (عملکرد مرجع) می‌شوند.
  2. مدل‌سازی: در این مرحله، یک مدل یادگیری عمیق یا شبکه عصبی معمولاً به عنوان شبکه تقلیدی ایجاد می‌شود. این شبکه به وضعیت فعلی عامل ورودی داده شده و سعی می‌کند عملی را تولید کند که به عملکرد مرجع نزدیک باشد.
  3. آموزش شبکه تقلیدی: شبکه تقلیدی با استفاده از داده‌های آموزشی، بهبود عملکرد خود را ادامه می‌دهد. این بهبود معمولاً توسط الگوریتم‌های یادگیری نظارت می‌شود.
  4. ارزیابی عملکرد: پس از آموزش، عامل یادگیرنده باید عملکرد خود را ارزیابی کند. این ارزیابی معمولاً با مقایسه عملکرد عامل یادگیرنده با عملکرد مرجع انجام می‌شود.
  5. تنظیم و بهبود: اگر عملکرد عامل یادگیرنده هنوز نسبت به مرجع کافی نباشد، ممکن است مراحل 2 تا 4 چندین بار تکرار شود تا عامل به تقلید بهتری از مرجع برسد.
- این روش به عنوان یکی از روش‌های ابتدایی در یادگیری تقویتی به کار می‌رود و به ویژه در مواردی که مشکل یادگیری تقویتی از ابتدا بدون مرجع مشخص و دقیق باشد، مفید است.
- در این پروژه تخصصی وجود ندارد که از او نحوه عملکرد مطلوب را جویا بشیم بنابراین این نوع یادگیری مطلوب نیست و حذف می‌شود

### 3-2-4 یادگیری تقویتی عمیق

- یادگیری تقویتی عمیق<sup>32</sup> یک شاخه از یادگیری ماشین است که به ترکیب دو عنصر اصلی، یعنی یادگیری تقویتی و شبکه‌های عصبی عمیق می‌پردازد. این روش برای حل مسائلی که تصمیم‌های توالی بر اساس عوامل متأثر از تصمیم‌ها نیاز دارند، مورد استفاده قرار می‌گیرد.
- شبکه‌های عصبی عمیق<sup>33</sup> نیز برای تقریب تابع ارزش عمل‌ها و تصمیم‌گیری عامل در یادگیری تقویتی استفاده می‌شوند. این شبکه‌ها به عنوان تقریب‌گرهای توابع ارزش عمل‌ها (Q-Values) عمل می‌کنند و عامل با استفاده از آن‌ها می‌تواند بهترین عمل را در هر وضعیت انتخاب کند.

---

<sup>32</sup> Deep Reinforcement Learning

<sup>33</sup> Deep Neural Networks

یکی از مطرح‌ترین الگوریتم‌های یادگیری تقویتی عمیق، الگوریتم شبکه Q عمیق است که توسط Google DeepMind معرفی شد و با استفاده از یک شبکه عصبی عمیق، عامل را آموزش می‌دهد تا عمل‌های بهتری انجام دهد. همچنین، الگوریتم‌هایی مانند عملگر-منتقد نیز برای یادگیری تقویتی عمیق مورد استفاده قرار می‌گیرند. [8]

یادگیری تقویتی عمیق در بسیاری از حوزه‌ها مانند بازی‌های ویدئویی، رباتیک، مدیریت منابع، بهینه‌سازی مالی، و زمینه‌های دیگر به کار می‌رود و به عنوان یکی از روش‌های پیشرو در یادگیری ماشین به شمار می‌آید.

### 3-3 استراتژی‌های کاوش

در یادگیری تقویتی، کاوش<sup>34</sup> و بهره‌برداری<sup>35</sup> دو مفهوم مهمی هستند که در فرآیند تصمیم‌گیری و یادگیری تقویتی نقش دارند. این دو مفهوم به شکل متعادلی با یکدیگر ترکیب می‌شوند تا بهترین عملکرد را به دست آورند. در ادامه به توضیح هر یک از این مفاهیم می‌پردازیم:

#### 1. کاوش

- کاوش به معنای انجام اقداماتی است که به شناخت محیط یادگیری و کشف اطلاعات جدید کمک می‌کند. در یادگیری تقویتی، انجام کاوش به منظور یافتن تجربیات جدید و تخمین ارزش‌های مختلف عملکردها استفاده می‌شود.
- کاوش می‌تواند با انتخاب عملکردهای تصادفی یا آزمون عملکردهای جدید انجام شود. این کار به افزایش دانش مدل در مورد محیط و اطلاعات مربوط به آن کمک می‌کند.

#### 2. بهره‌برداری

- بهره‌برداری به معنای انجام اقداماتی است که بر اساس اطلاعاتی که تا کنون جمع آوری شده است، بهترین عملکرد را انتخاب می‌کند. در واقعیت، این به انجام عملکردهایی که تا کنون به عنوان موثرترین شناخته شده‌اند و باعث کسب مزیت‌های بیشتر می‌شوند، اشاره دارد.

---

<sup>34</sup> Exploration

<sup>35</sup> Exploitation

- بهره‌برداری به منظور بهره‌گیری از تجربیات قبلی و بهبود عملکرد با استفاده از اطلاعات به دست آمده از مصالحه انجام می‌شود.

تبادل میان کاوش و بهره‌برداری در یادگیری تقویتی مهم است. اگر تنها مصالحه انجام شود، ممکن است مدل به طور مداوم عملکردهای تصادفی انجام دهد و عملکرد بهبود نیابد. از طرف دیگر، اگر تنها بهره‌برداری صورت گیرد، ممکن است اطلاعات جدید در مورد محیط جمع آوری نشود و عملکرد به حداکثر خود نرسد. بنابراین، تبادل مناسب بین این دو عنصر باید در الگوریتم‌های یادگیری تقویتی حفظ شود تا بهترین عملکرد ممکن به دست آید. [9]

اپسیلون<sup>36</sup> در الگوریتم‌های یادگیری تقویتی به عنوان یک پارامتر مهم و تصمیم‌گیری در روشهای تصادفی مورد استفاده قرار می‌گیرد. اپسیلون معمولاً به عنوان مقداری بین صفر و یک تعریف می‌شود و نمایانگر نسبتی از زمان‌هاست که عامل یادگیری در انتخاب عمل تصادفی عمل می‌کند به جای انتخاب بهترین عمل بر اساس استراتژی یادگیری تصمیم‌گیری یا همان سیاست خود.

به عبارت دیگر، اگر مقدار اپسیلون برابر با صفر باشد، عامل همیشه عمل بهترین تخمین‌گر برای عمل بهره‌برداری را انجام می‌دهد و هیچ عمل تصادفی انجام نمی‌دهد. اما اگر اپسیلون برابر با یک باشد، عامل همیشه تصادفی عمل می‌کند و هیچ‌گاه به عمل بهینه تخمین‌گر نمی‌پردازد و کاوش می‌کند. در حالت عمومی، اپسیلون می‌تواند یک مقدار میانی بین صفر و یک داشته باشد، که عامل به تصادف عمل می‌کند با احتمال اپسیلون و به تمام تصمیمات اشتباه از تصمیمات تخمین‌گر بهینه دوری می‌کند.

استفاده از اپسیلون در الگوریتم‌های یادگیری تقویتی به عامل این امکان را می‌دهد که در طول زمان از تجربیات خود بیاموزد و به بهبود عملکرد خود برسد، در حالی که همچنان از تصمیمات بهینه‌ای که تاکنون یاد گرفته است، استفاده کند. این تبادل بین اکتشاف و بهره‌برداری به عامل اجازه می‌دهد که به دنبال جستجوی بهترین راه حل‌ها باشد و به دست آوردن تجربیات جدید را نیز تشویق کند.

### 3-3-1 اپسیلون-حریص

---

<sup>36</sup> Epsilon

الگوریتم اپسیلون-حریص<sup>37</sup> یکی از روش‌های استفاده شده در مسائل تقسیم منابع برای تجربه کردن و تست کردن چندین گزینه است. این الگوریتم به صورت معمول در مسائلی مورد استفاده قرار می‌گیرد که می‌خواهیم بین چندین گزینه یا عمل مختلف انتخاب کنیم و به طور متوسط بهترین گزینه را پیدا کنیم.

### 3-3-2 مرز بالای اعتماد<sup>38</sup>

این مفهوم به تعیین یک سیاست تصمیم‌گیری برای یک عامل یادگیری تقویتی اشاره دارد که برای انتخاب اعمال (actions) مختلف در یک محیط تعاملی استفاده می‌شود.

الگوریتم مرز بالای اعتماد معمولاً در مسائلی به کار می‌رود که به عامل امکان برخورد با شرایط مختلف در محیط را می‌دهند و او نیاز به بیشترین بهره‌برداری از اعمال بازدهی بالا دارد.

مرز بالای اطمینان، برای هر عمل، یک مقدار مرز بالای اطمینان بر اساس داده‌های تاکنون جمع‌آوری شده محاسبه می‌شود. این مقدار نمایانگر اطمینان ما از بازدهی این عمل است و معمولاً با استفاده از اندازه نمونه‌ها و مقدارهای آماری مرتبط محاسبه می‌شود.

با توجه به مقدار مرز بالای اطمینان برای هر عمل، عامل یادگیری تقویتی انتخاب می‌کند که کدام عمل را انجام دهد. این استراتژی کمک می‌کند تا عامل به صورت تدریجی عمل‌های با بالاترین بازدهی را ترجیح دهد و در عین حال اطمینان از صحت تخمین‌هایش را حفظ کند.

الگوریتم مرز بالای اعتماد به طور گسترده در مسائل یادگیری تقویتی مورد استفاده قرار می‌گیرد و بهبود عملکرد عوامل در محیط‌های تعاملی را تسهیل می‌کند.

### 3-3-3 نمونه‌برداری تامپسون<sup>39</sup>

نمونه‌برداری تامپسون یک الگوریتم معروف در حوزه یادگیری تقویتی است که برای مسائل تصمیم‌گیری تعاملی با استفاده از اصول بیزی مورد استفاده قرار می‌گیرد. این الگوریتم به عامل یادگیری تقویتی امکان می‌دهد تا بر اساس تخمین توزیع احتمالی مقادیر بازدهی عمل‌ها تصمیم بگیرد.

---

<sup>37</sup> Epsilon Greedy

<sup>38</sup> Upper Confidence Bound (UCB)

<sup>39</sup> Thompson Sampling



عملکرد الگوریتم نمونه‌برداری تامپسون به صورت زیر است:

1. هر عمل ممکن در مسئله با یک توزیع احتمالی مرتبط است که نشان‌دهنده توزیع بازدهی آن عمل می‌باشد.
  2. الگوریتم نمونه‌برداری تامپسون برای هر عمل یک تخمین از توزیع احتمالی بازدهی آن عمل می‌سازد. این تخمین با استفاده از تحلیل بیزی و استفاده از تاریخچه‌ی بازدهی‌ها به‌روز می‌شود.
  3. در هر مرحله، الگوریتم یک نمونه تصادفی از توزیع احتمالی بازدهی هر عمل انتخاب می‌کند. این نمونه‌برداری تصادفی باعث می‌شود که عملی با بیشترین احتمال بازدهی بالا انتخاب شود.
  4. عملی که در مرحله 3 انتخاب شده است، انجام می‌شود و بازدهی آن به عامل یادگیری اطلاع داده می‌شود.
  5. تخمین‌های بیزی برای توزیع‌های بازدهی عمل‌ها بر اساس داده‌های جدید به‌روزرسانی می‌شود.
- این چرخه تکرار می‌شود و الگوریتم تامپسون ادامه می‌دهد تا تاخیر زمانی مشخصی یا تا رسیدن به یک تعداد معین از مراحل. این الگوریتم به عامل این امکان را می‌دهد که به طور آگاهانه و با توجه به تخمین‌های بیزی، تصمیم‌های بهینه‌تری در مورد انتخاب عمل‌ها بگیرد.
- الگوریتم تامپسون معمولاً در مسائلی که توزیع بازدهی عمل‌ها تغییر پذیر است یا در مسائلی که نیاز به کاوش و بهره‌برداری همزمان دارید مورد استفاده قرار می‌گیرد و به عنوان یکی از الگوریتم‌های اصولی یادگیری تقویتی به شمار می‌آید.

## 4-3 جمع‌بندی

الگوریتم‌های یادگیری تقویتی را می‌توان به طور عمده به دسته‌های یادگیری مدل‌دار و یادگیری بدون مدل تقسیم‌بندی کرد. الگوریتم‌های مدل‌دار سعی در ترسیم یک مدل از محیطی که با آن در تعامل هستند و پویایی آن‌ها بپردازند در حالی که در سیستم استخراج سرمایه‌گذاری قیمت ارز خروجی و همچنین نرخ سرمایه‌گذاری کمتر دارای مدل مشخصی هستند. بنابراین استفاده از یک مدل یادگیری بدون مدل راهکار مناسب‌تری است. همچنین در این مسئله بایستی بتوانیم که با مدل مناسبی بتوانیم که مصالحه کاوش و بهره‌برداری را حل نماییم. اگر که راهکارهای موجود نگاهی بیندازیم می‌بینیم که الگوریتم نمونه‌برداری تامپسون بهترین راهکار برای این موضوع است زیرا اگر که از سایر الگوریتم‌ها استفاده کنیم ممکن است که عامل رفتارش بعد از مدتی ثابت شود ولی قیمت و میزان

## فصل سوم: روش پیشنهادی و نتیجه‌گیری

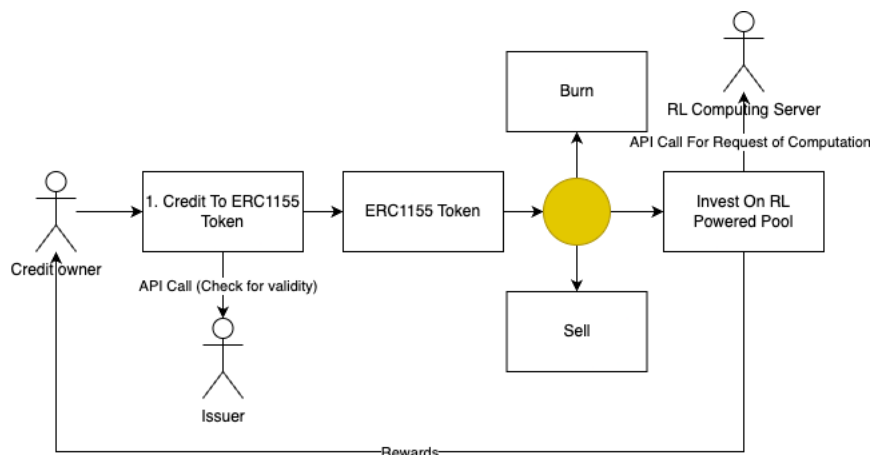
سرمایه‌گذاری این‌چنین نیست. پس نیاز داریم که با راهکاری که هم‌زمان امکان کاوش و بهره‌برداری را ارائه می‌دهد پیش برویم.

## فصل چهارم روش پیشنهادی و روش پیاده‌سازی

## 4-1 مقدمه

در این فصل به نحوه پیاده‌سازی پل تبدیل اعتبارات کربنی به رمزارزها و همچنین نحوه پیاده‌سازی استخر تبدیل اعتبارات کربنی رمزارزی به دست آمده به رمزارزهای با نقدینگی بالا می‌پردازیم. همچنین بایستی بیان کنیم که چگونه در این سیستم به ایمنی لازم دسترسی پیدا کرده‌ایم. ایمنی در قراردادهای هوشمند در بستر بلاکچین دارای اهمیت بالایی است زیرا در صورتی که یک قرارداد در محیط بلاکچین مستقر شود، دیگر امکان تغییر آن وجود ندارد و همچنین این قراردادها با پول سر و کار دارند، اگر که امکان حمله به آن‌ها وجود داشته باشد دارایی از دست خواهد رفت. همچنین لازم به ذکر است که یادگیری تقویتی بار پردازشی زیادی لازم دارد و همچنین لازم به حافظه زیادی است که سیستم بلاکچین این امکان را به ذاتا در اختیار ما نمی‌گذارد.

## 4-2 نحوه پیاده‌سازی بخش تبدیل اعتبار کربن به رمزارز



- فرایند سرمایه‌گذاری 4 تصویر

فرایند سرمایه‌گذاری که در تصویر ۴ به نمایش گذاشته شده است مطابق زیر است:

1. ابتدا صاحب اعتبار کربن باستی در سازمانی که از آن اعتبار کربن خود را گرفته درخواست تبدیل آن به نوع رمزارزی را بدهد.
2. در صورت پذیرش سازمان نام‌برده کاربر از طریق قرارداد هوشمند یک توکن ERC1155 خالی تولید نموده و با استفاده از آن درخواستی به سازمان تولید کننده اعتبار می‌زند و به این صورت از طریق یک تماس API از طریق یک Oracle می‌توان اعتبار را به صورت رمزارزی با قالب توکن ERC1155 در آورد.

3. صاحب اعتبار می‌تواند با استفاده از توکن ERC1155 که معادل فیزیکی اعتبار کربن است، اقدامات متنوعی بکند از جمله:

a. فروش

b. سوزاندن

c. سرمایه‌گذاری

4. در صورتی که راه آخر یعنی سرمایه‌گذاری را انتخاب کند، با استفاده از یک تماس Oracle به سروری مجهز به یادگیرنده تقویتی سرور از سرمایه‌گذاری شخص سرمایه‌گذار در استخر مربوطه مطلع می‌شود. پس در نتیجه شروع به محاسبه تدریجی نرخ سود او می‌کند.

5. کاربر هر موقع که بخواهد می‌تواند با درخواست از طریق قرارداد هوشمند توکن‌های قابل تعویض ERC20 به ازای مدت سرمایه‌گذاری در صندوق دریافت کند.

### **3-4 توکن‌های ERC1155 با یکدیگر متفاوت هستند**

یکی از مزایای پروتکل ERC1155 این است که هم‌زمان می‌تواند هم قابل تعویض و غیرقابل تعویض باشد. در این پروژه توکن‌های ERC1155 از ۴ نوع متفاوت هستند که به صورت زیر می‌باشد.

1. نوع استاندارد: سازمان صادر کننده استاندارد

2. نوع جبران کربن: نحوه مقابله با گرمایش زمین مثلاً: تولید انرژی پاک، ممانعت از جنگل‌زدایی، جنگل‌زایی

3. حجم جبران کربن: میزان کربن جبران شده از سطح زمین

4. سال صدور

### **4-4 راهکار یادگیری تقویتی برای تبدیل ERC-1155 به ERC-20**

تقاضا و عرضه دو مفهوم کلیدی در اقتصاد هستند که نقش مهمی در تعیین قیمت‌ها و تعاملات اقتصادی ایفا می‌کنند. این دو عنصر به طور مستقیم با یکدیگر مرتبط هستند و تأثیرات متقابلی بر روی اقتصاد دارند. در زیر به توضیح هر یک از این مفاهیم می‌پردازیم:

1. تقاضا: تقاضا به میزان کالا یا خدماتی اشاره دارد که افراد یا سازمان‌ها در بازار خریداری می‌کنند. تقاضا معمولاً به عواملی مانند قیمت، درآمد، ترجیحات مصرف‌کنندگان، توقعات و شرایط اقتصادی بستگی دارد. افزایش تقاضا ممکن است به افزایش قیمت‌ها منجر شود، به عنوان مثال وقتی که تقاضا بیشتر از عرضه است. تقاضا می‌تواند بر اساس تغییرات فصلی، ترتیب محصولات جدید و ابتکارات صنعتی نیز تغییر کند.

2. عرضه: عرضه به تعداد کالا یا خدماتی اشاره دارد که توسط تولیدکنندگان یا فروشندگان در بازار قرار داده می‌شود. عرضه نیز به عواملی مانند هزینه تولید، تکنولوژی، تغییرات در منابع، واکنش به تغییرات قیمت و سایر عوامل اقتصادی بستگی دارد. افزایش عرضه ممکن است منجر به کاهش قیمت‌ها شود، به عنوان مثال وقتی که عرضه بیشتر از تقاضا است.

تبادل بین تقاضا و عرضه در بازار می‌تواند به تنظیم قیمت‌ها و مقدار تولید مناسب منجر شود. اگر تقاضا بیشتر از عرضه باشد، قیمت‌ها ممکن است افزایش پیدا کنند، و اگر عرضه بیشتر از تقاضا باشد، قیمت‌ها ممکن است کاهش یابد. این فرآیند تعیین قیمت‌ها را به بازار و نیروهای اقتصادی ترک می‌کند. تغییرات در تقاضا و عرضه می‌توانند به دلایل مختلفی از جمله تغییرات اقتصادی، تغییرات در سیاست‌های دولتی و تغییرات در تکنولوژی رخ دهند و تأثیرات گسترده‌ای بر اقتصاد داشته باشند.

یکی از عوامل موثر بر عرضه یک رمزارز، میزان تولید آن است. به طور مثال بیت‌کوین تقریباً هر چهار سال یک بار، یا به عبارت دیگر هر ۲۱۰ هزار بلاک یک‌بار میزان عرضه بیت‌کوین‌ها را نصف می‌کند. این منجر به ایجاد یک تصاعد هندسی در میزان عرضه این رمزارز شده که منجر به این می‌شود که میزان محدودی از این ارز موجود باشد و دچار تورم نشود.

در اینجا نیز ارزش رمزارز ERC-20 بسته به میزان تولید آن می‌تواند متفاوت باشد. هرچه بیشتر این ارز تولید شود کم‌یابی آن کمتر شده و از ارزش آن کاسته می‌شود. هرچه که این ارز کمتر تولید شود نیز برعکس.

حال اگر که جای اینکه میزان معینی از رمز ارز ERC-20 خلق کنیم اگر که آن‌ها را در یک استخر سرمایه‌گذاری قرار دهیم که با فرکانس مشخصی به دارندۀ ERC-1155 معادل اعتبار کربن به آن‌ها ERC-20 می‌دهد، آیا می‌توان هرچقدر که لازم است فرکانس مذکور را کمتر کرد؟ به طور حتم خیر زیرا در این صورت دیگر ارز ERC-20 ای تولید نمی‌شود که بخواهد از ارزش آن کم شود. و به طبع اگر که میزان عرضه آن نیز کمتر شود میزان ارز که به دست سرمایه‌گذار می‌رسد ممکن است که مورد پسند او نباشد پس لازم است که در این جا یک مصالحه صورت بگیرد.

یادگیری تقویتی می‌تواند در اینجا مورد استفاده قرار بگیرد. حال به بیان نحوه استفاده از یادگیری تقویتی در این مسئله می‌پردازیم.

یادگیری تقویتی در این جا به صورت زیر تعریف می‌شود:

1. وضعیت: میزان قیمت ارز ERC-20 خروجی و همچنین میزان سرمایه‌گذاری کاربران
2. پاداش: افزایش ارز ERC-20 معادل پاداش مثبت و کاهش آن نیز معادل پاداش منفی است. افزایش میزان سرمایه‌گذاری در استخر نیز پاداش مثبت و کاهش آن نیز پاداش منفی است.
3. عمل: عامل می‌تواند با افزایش یا کاهش فرکانس تولید ERC-20 با محیط تعامل کند.
4. عامل: موجودیتی که میزان تولید ERC-20 را تعیین و با توجه به وضعیت موجود بهینه‌ترین عمل را تشخیص می‌دهد.

## 4-5 چرا یادگیری تقویتی؟

در بخش قبل گفتیم که عرضه و تقاضا یکی از عوامل با ارزش شدن یک کالا می‌باشد. اما میزان موجودی تنها عامل تاثیرگذار بر قیمت نیست. به طور کمی دقیق‌تر میزان موجودی نقدی است که میزان قیمت را تعیین می‌کند. این عامل تنها در دست میزان تولید کالا نیست. بلکه عملکرد مردم در قبال آن کالا نیز بر میزان قیمت آن تاثیرگذار است. اگر که میزان قیمت یک کالا را به فرض محدود به میزان موجودی آن به طور کلی نکنیم، هر چه قیمت آن بالاتر رود تمایل به کسب آن کالا نیز بیشتر می‌شود. در مسئله ما اگر که به دلیلی غیر از عرضه و تقاضا ناشی از میزان تولید، میزان سرمایه‌گذاری جهت کسب رمزارز خروجی استخر سرمایه‌گذاری بیشتر شود، این برای صاحب استخر منفعت دارد. افزایش قیمت نیز می‌تواند ناشی از کاهش موجودیت ناشی از کمتر تولید کردن رمزارز قابل تعویض باشد که خود می‌تواند عاملی برای افزایش سرمایه‌گذاری شود. پس به طور کلی میزان قیمت و میزان سرمایه‌گذاری می‌توانند به طور مستقل عاملی برای کاهش یا افزایش میزان تولید رمزارز خروجی استخر باشند. پس می‌توان به طور خلاصه در جدول زیر گفت:

جدول 2- عوامل موثر بر میزان تولید رمزارز خروجی استخر

عوامل افزایش میزان تولید	عوامل کاهش میزان تولید
افزایش تقاضا به دلیل افزایش قیمت مستقل از میزان تولید	افزایش قیمت به دلیل کاهش میزان تولید

افزایش تقاضا به دلیل افزایش قیمت ناشی از کمتر شدن تولید	
---	--

لازم به ذکر است که در اینجا تغییر میزان تولید یعنی عمل عامل بر محیط مسئله یعنی میزان قیمت و سرمایه‌گذاری تاثیر گذار است. پس نیازمند نوعی هوشمندی داریم که بتواند از تاثیر اعمال خود بر محیط آگاه باشد.

## 4-6 انتخاب مدل یادگیرنده تقویتی

حال با توجه به تمامی این موارد می‌توان گفت که با انتخاب نوع مطلوب مدل یادگیری تقویتی می‌توان مسئله را حل نمود. با توجه به فصل قبل یادگیری تقویتی را می‌توان به دو دسته تقسیم کرد: یادگیری با مدل و یادگیری بدون مدل. در یادگیری با مدل، الگوریتم‌ها سعی می‌کنند یک مدل از محیطی که با آن تعامل دارند بسازند و با پویایی محیط مدل را به‌روز کنند. اما در یادگیری بدون مدل، سیستم‌ها در تصمیم‌گیری‌های خود از مدل خاصی استفاده نمی‌کنند و به جای آن، به صورت مستقیم با محیط تعامل می‌کنند. از این رو، در مواجهه با وظیفه‌هایی مانند پیش‌بینی قیمت ارز و تصمیم‌گیری در مورد فرکانس استخراج سرمایه‌گذاری، یادگیری بدون مدل رویکرد بهتری به نظر می‌آید. در این پروژه به دو نوع یادگیرنده بدون مدل می‌پردازیم و بیان می‌کنیم که هر یک چگونه عمل می‌کند و ما چگونه از یادگیرنده‌ها بهره می‌بریم.

### 1-4-6 یادگیرنده تقویتی شبکه یادگیری عمیق Q

این یک نوع یادگیرنده بدون مدل است که سعی در کسب یک سیاست برای نحوه عملکرد در شرایط مختلف بدهد. این یادگیرنده این یادگیرنده از ترکیب یادگیرنده Q و شبکه‌های عصبی عمیق به دست می‌آید.

#### 1-4-6-1 یادگیرنده Q

این الگوریتم، عامل از یک جدول به نام "Q-Table" استفاده می‌کند که ابعاد آن برابر با تعداد وضعیت‌ها و تعداد عمل‌ها در محیط است. هر خانه در این جدول دارای مقداری به نام Q-value است که نمایانگر ارزش ترکیب وضعیت و عمل است. عامل با انجام عملی در محیط و مشاهده پاداش‌های دریافتی، تجربی اقدامات خود را بهبود می‌دهد. از این تجربیات به منظور بهبود استراتژی انتخاب عمل‌ها استفاده می‌کند و مقادیر Q-value را در جدول به روز می‌کند. این الگوریتم به ویژه برای محیط‌های قطعی مناسب است، اما می‌توان آن را به محیط‌های غیرقطعی نیز اعمال کرد.



این عامل از رابطه زیر برای سیاست‌گذاری‌های خود استفاده می‌کند.

$$Q(s, a) = Q(s, a) + \alpha \times [R(s, a) + \gamma \times \max_{a'} Q(s', a') - Q(s, a)]$$

در این فرمول:

- $Q(s, a)$  : برای وضعیت  $s$  و عمل  $a$ .
- $\alpha$  (نرخ یادگیری) : یک پارامتر که نشان‌دهنده میزان اهمیت تجربیات جدید است و می‌تواند بین 0 و 1 باشد.
- $R(s, a)$  : پاداش دریافتی برای انجام عمل  $a$  در وضعیت  $s$ .
- $\gamma$  (عامل تخفیف) : یک پارامتر که نشان‌دهنده تاثیر پاداش‌های آینده است و می‌تواند بین 0 و 1 باشد.
- $\max(Q(s', a'))$  : بیشترین Q-value برای وضعیت بعدی  $s'$  و همه عمل‌های ممکن در آن وضعیت.
- $Q(s, a)$  : مقدار Q-value قبلی برای وضعیت  $s$  و عمل  $a$ .

عامل با استفاده از این فرمول مقدار Q-value را به روزرسانی کرده و به تدریج استراتژی بهینه‌ای برای تصمیم‌گیری در محیط را یاد می‌گیرد. این فرآیند ادامه پیدا می‌کند تا به مرور بهینه شود. [8]

#### 2-1-6-4 شبکه یادگیری عمیق

از این شبکه‌ها برای یادگیری جدول Q در شبکه نوروهای عصبی استفاده می‌شود. یعنی به طور تدریجی مقادیر Q-Value کسب شده و یادگیرنده عمیق آن‌ها را به مرور یاد می‌گیرد.

#### 4-6-1-3 نکاتی در خصوص پیاده‌سازی

لازم به ذکر است که در این مسئله جدول Q به طور ضمنی توسط یادگیرنده عمیق یادگرفته می‌شود و ما هرگاه تعداد داده‌هایمان از چهار مورد وضعیت فعلی، پاداش، وضعیت بعدی، اتمام به عدد ۶۴ رسید همه این داده‌های را در قالب یک batch به یادگیرنده عصبی عمیق می‌دهیم تا یادگیری جدول و سیاست‌های ناشی از آن صورت گیرد.

#### 4-6-2 یادگیرنده عمیق SARSA

این یادگیرنده تماماً معادل یادگیرنده تقویتی Q است با این تفاوت که به جای پیدا کردن بهترین عمل ممکن بعد از عمل  $a$  از سیاست تا کنون یادگرفته شده برای تعیین آن استفاده می‌کند. یعنی رابطه توصیف‌گر آن به صورت زیر می‌شود.

$$Q(s,a) = Q(s,a) + \alpha \times [R(s,a) + \gamma \times Q(s',a') - Q(s,a)]$$

## 4-7 تعیین تابع پاداش

همان‌طور که قبل‌تر بیان شد، پاداش و همچنین عِقاب یک یادگیرنده تقویتی با استفاده از میزان افزایش و کاهش میزان قیمت و سرمایه‌گذاری تعریف می‌شود.

حال برای هر یک نیاز به دو تابع داریم تا معیاری عددی برای هریک ارائه کند.

### 4-7-1 تابع قیمت

برای تعیین تاثیر قیمت در پاداش از مشتق زمانی آن یعنی آخرین قیمت منهای یک مانده به آخرین قیمت یا به عبارت دیگر از عبارت زیر استفاده کردیم.

$$price(t) - price(t - time\ unit)$$

### 4-7-2 تابع سرمایه‌گذاری

برای تعیین میزان تاثیر سرمایه‌گذاری کمی متفاوت عمل کردیم. به این صورت که مشروط به آنکه دو سرمایه‌گذاری صورت گرفته اخیر فاصله‌ای کمتر از واحد زمانی داشته باشند، پاداش معادل اختلاف آن‌ها باشد. یعنی:

$$investment(last) - investmnet(last - 1) \quad if \quad timebetween them < time\ unit$$

در نهایت میزان پاداش نهایی جمع وزن‌داری از آن‌ها است. که وزن هر یک در صورت تغییر می‌تواند منجر به تغییر مسئله شود. پس با توجه به نیازمندی‌های خاص مد نظر سرمایه‌گذار می‌توان آن‌ها تعیین کرد.

## 4-8 مصالحه کاوش و بهره‌برداری

در این مسئله گاه ممکن است، عموماً رفتار انسان‌ها امری ثابَل پیش‌بینی نیست و پایه چنین تشخیص الگوهایی در بازارهای مالی و ارزی در امری روان‌شناسی است که مردم در شرایط یکسان اعمال یکسانی دارند. اگر چه ما در این پروژه سعی در این داریم که از الگوهای پیشین با استفاده از یادگیرنده تقویتی عملکرد مردم در آینده را پیش‌بینی نموده و طبق آن عمل کنیم، ولی لزوماً داده‌هایی که طبق آن عامل یادگیری را انجام داده حاکی از تمام شرایطی نیست که مردم در آن تصمیم‌گیری خود را انجام داده‌اند. این اصل در امر تحلیل تکنیکال نیز به همین صورت است و مردم ممکن است طبق آنچه که تحلیل می‌گویند عمل نکنند. بلکه عواملی بیرونی به جز روندی که قیمت تا کنون طی کرده مانند تاثیر سیاست‌های حکومت نیز در این موارد تاثیر گذار باشد، که به اصطلاح به آن

عوامل بنیادی<sup>40</sup> می‌گویند. پس ممکن است یادگیری که تا کنون صورت گرفته نادرست باشد و عامل بایستی راهکاری برای تغییر سیاست‌های خود پس از مدتی یادگیری سیاست داشته باشد. این امر نیازمند کاوش کردن پس مدتی بهره‌برداری از آن چه تا کنون یاد گرفته شده است. پس در این مسئله ابتدا نیاز به انجام کاوش تا جای ممکن و سپس انجام عمل بهره‌برداری تا جایی که عملکرد مطلوب است هستیم. انجام کاوش و سپس بهره‌برداری از آن چیزی است که الگوریتم اپسیلون-حریص ارائه می‌دهد و انجام عمل مکاشفه در حین انجام بهره‌برداری چیزی است که الگوریتم تامپسون ارائه می‌دهد. از طرفی نیز مطلوب است که تناظری بین عملکرد مطلوب و غیر مطلوب عامل و همچنین میزان مکاشفه و بهره‌برداری داشته باشیم. برای این منظور بایستی توزیعی به تناسب عملکرد مطلوب و غیر مطلوب عامل داشته باشیم. توزیع بتا برای این مورد می‌تواند توزیع مناسبی باشد. در اینجا یک الگوریتم هیبریدی از نمونه‌برداری تامپسون و الگوریتم اپسیلون-گریدی ارائه می‌دهیم که منجر به عملکرد بهینه شود:

1. الگوریتم اپسیلون-حریص تا جایی که دیگر تغییر نکند ادامه می‌یابد. یعنی به حد پایینی از پیش تعیین شده برسد اجرا می‌شود.

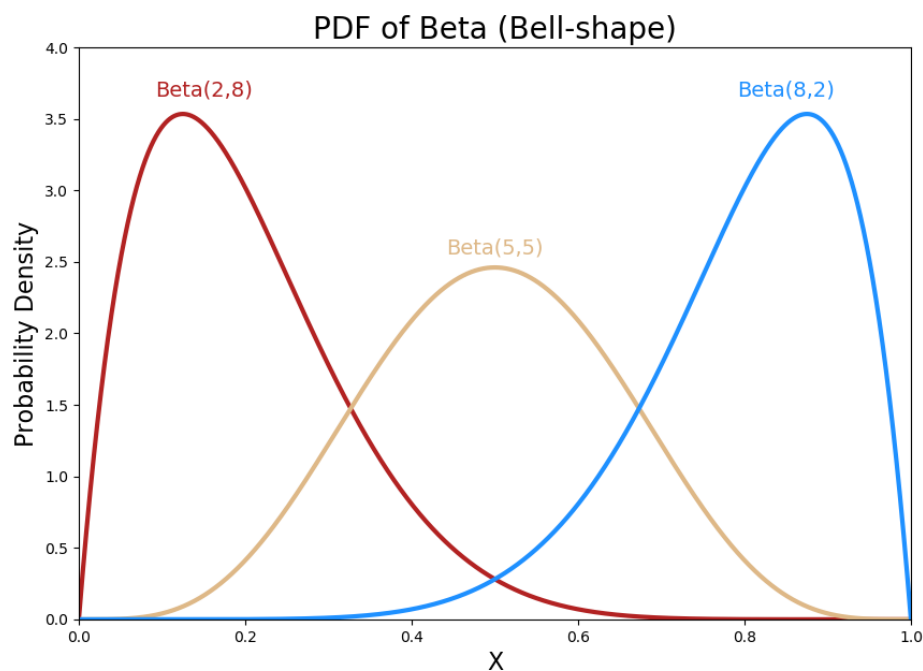
2. سپس الگوریتم نمونه برداری تامپسون اجرا می‌شود. به این صورت که علامت آخرین پاداش منهای پاداش قبلی حاکی از عملکرد خوب یا بد است. میزان عملکرد خوب و بد یک جا ذخیره می‌شود. سپس با استفاده از آن‌ها هنگام تصمیم در مورد کاوش یا بهره‌برداری یک متغیر تصادفی از توزیع بتا تولید می‌شود. این توزیع دارای فرمول چگالی زیر و شکل زیر نیز برخی از مقادیر آلفا و بتا را برای آن نشان می‌دهد.

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

در این فرمول در مسئله ما  $a$  معادل عملکرد خوب و  $b$  معادل عملکرد بد است. همچنین  $\Gamma(n)$  تابع گاما یا همان  $factoriel(n-1)$  برای اعداد طبیعی است.

---

<sup>40</sup> Fundamental



تصویر 5- توزیع بتا منبع وبسایت Towards Data Science

## 4-9 معماری شبکه عصبی

یادگیرنده‌های عمیق Q و SARSA همانگونه که توضیح داده شد از یک شبکه عصبی برای یادگیری سیاست‌های خود بهره می‌برند. این شبکه عصبی در این پروژه دارای ۴ لایه است که در جدول زیر آمده است.

Input layer	Hidden Layer One	Hidden Layer Two	Output Layer
Number of 2 neurons	Number of 64 neurons	Number of 64 neurons	Number of 21 neurons

1. در لایه اول دو نورون وجود دارد که حاکی از وضعیت فعلی یعنی قیمت و پاداش می‌باشد.

2. در لایه دوم یا لایه پنهان اول ۶۴ نورون وجود دارد که دارای فعال‌ساز<sup>41</sup> relu می‌باشند هست.

<sup>41</sup> Activation function

3. در لایه سوم یا لایه پنهان دوم نیز ۶۴ نورون وجود دارد که دارای فعال‌ساز  $\text{relu}$  می‌باشند هست.

4. در لایه آخر ۲۱ نورون هستند که حاکی از بالا تا پایین حاکی از عمل کاهش فرکانس از -۱۰ تا افزایش فرکانس تا ۱۰ می‌باشند که عمل انتخابی آن است که بیشترین مقدار خروجی را دارد.

## 4-10 شبیه‌سازی

از آن‌جا که اصل کار یادگیرنده‌های تقویتی در محیط تعریف می‌شود استفاده از داده‌های از پیش دست یافته شده برای آن شدنی نیست. همچنین تولید داده‌هایی برای قیمت و میزان سرمایه‌گذاری، تاثیر عمل خود عامل را مد نظر نمی‌گیرد.

راهکار چیست؟ پیاده‌سازی یک محیط شبیه‌سازی که در آن تا حدودی عملکرد مردم با توجه به عملکرد عامل تغییر می‌کند.

برای این منظور ابتدا چند نمودار قیمت که ۱۳ تا از آن‌ها از قواعد الگوهای کلاسیک تحلیل تکنیکال پیروی کرده و ۱۲ تا از آن‌ها از قواعد خاصی پیروی نمی‌کنند تولید می‌کنیم. نمودار میزان فرکانس سرمایه‌گذاری را رشد حرکتی نمایی قیمت در ابتدا فرض می‌کنیم. دلیل انتخاب ۱۳ نمودار از روی الگوهای کلاسیک تحلیل تکنیکال این است که قیمت رمزارز خروجی مانند سایر رمزارزها احتمالاً با توجه به عملکرد عموم مردم (در صرافی‌های رمزارزی) در شرایط یکسان از الگوهای پیشینی که در بازارهای ارزی هست پیروی کند. همچنین همواره نیز رفتار مردم الگوپذیر نیست پس نیازمند ترکیبی از نمودارهای قیمت هستیم که تعدادی از الگوهای پیشین پیروی نموده و تعدادی نیز از الگویی پیروی نکنند و تنها الگویی رندوم داشته باشند.

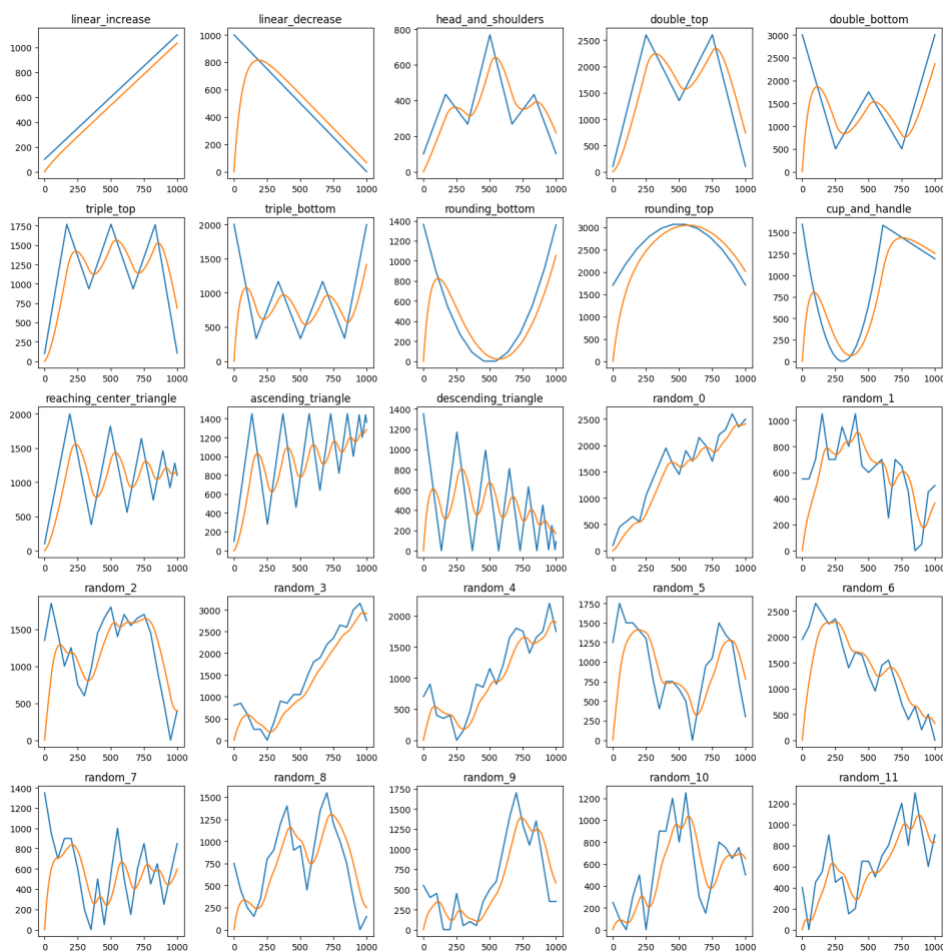
نحوه به دست آوردن میزان سرمایه‌گذاری از روی میزان قیمت نیز بدون در نظر گرفتن تاثیر عمل عامل بر این دو نیز خود یک سوال است که باید به آن جواب دارد. می‌توان گفت که مردم با کمی تاخیر با توجه به آنچه که روند تا کنون بوده به دنبال کسب بیشتر آن ارز هستند، دلیل آن نیز این است که تا یک روند توسط مردم تشخیص داده شود کمی زمان می‌برد و مردم در شروع هر روند نمی‌توانند آن روند را در جا تشخیص دهند، بلکه با اختلافی آن را متوجه می‌شوند. برای دستیابی به چنین اختلافی در میزان سرمایه‌گذاری و قیمت می‌توان میزان سرمایه‌گذاری را معادل میانگین نمایی حرکت قیمت دانست. لازم به ذکر است که این یک فرض برای انجام این شبیه‌سازی است. نبود نمونه پیشین از استخراج سرمایه‌گذاری رمزارزی و علی‌الخصوص استخراج سرمایه‌گذاری رمزارزهای اعتبار کربن باعث می‌شود که دیدگاه دقیقی از آنچه که مردم عمل می‌کنند نداشته باشیم ولی می‌توان با آنچه که در گذشته بر قیمت رمزارزهای دیگر آمده و عملکرد مردم در بازارهای مالی دیگر این فرض‌ها را برای

این استخرهای سرمایه‌گذاری داشت. رابطه‌ای که با آن میانگین نمایی حرکتی برای یک روند تعریف می‌شود به صورت زیر است:

$$T(n) = \alpha \times price(n) + (1 - \alpha) \times T(n - 1)$$

که در این رابطه  $T$  معادل میانگین نمایی حرکتی می‌باشد و  $\alpha$  معادل 0.02 در نظر گرفته شده است.

پس نمودارها به صورت زیر می‌شوند که در آن‌ها آبی قیمت و زرد میزان سرمایه‌گذاری است.



این‌ها البته داده‌های خام اولیه می‌باشند. سپس در ادامه با یک احتمال و با توجه به اینکه قیمت تناسب عکس با

میزان تولید داشته و میزان سرمایه‌گذاری تناسب مستقیم با قیمت داشته و میزان سرمایه‌گذاری تناسب مستقیم با

میزان تولید داشته با احتمالی این روابط را حین اجرای شبیه‌سازی به قیمت و میزان سرمایه‌گذاری دخیل می‌کنیم که

این احتمالا در جدول زیر آمده است.

0.02	احتمال تغییر قیمت با توجه به میزان تولید (عکس)
۰.۰۳	احتمال تغییر میزان سرمایه‌گذاری با توجه به میزان تولید (مستقیم)
۰.۰۲	احتمال تغییر میزان میزان سرمایه‌گذاری با توجه به تغییرات قیمت (مستقیم)

انتخاب مناسب هر یک از احتمالات گفته شده به نزدیکی بیشتر شبیه‌سازی به واقعیت کمک می‌کند. احتمالات گفته شده با توجه به میزان پایداری تغییرات عدم ایجاد اختلاف زیاد با آنچه که الگوی اصلی است و با آزمون و خطاهای متوالی انتخاب شده و به دست آمده است. فرض شده که احتمال تغییر قیمت به با توجه به تولید احتمالی یکسانی با تغییر سرمایه‌گذاری با قیمت دارد و این احتمال کمتر از احتمال تغییر میزان سرمایه‌گذاری با توجه به میزان تولید است. تضمین نیست که این اعداد با واقعیت تطابق داشته باشند ولی برای قصد شبیه‌سازی ناچار به انتخاب این مقادیر هستیم.

شبیه‌سازی معادل اجرای عامل در شرایطی است که قیمت و سرمایه‌گذاری از پویایی گفته شده و همچنین داده‌های خام گفته شده اجرا می‌شود و با توجه به پاداش‌ها یادگیری‌اش صورت می‌گیرد. همچنین پاداش و عقاب نیز مطابق آنچه که گفته شد صورت می‌گیرد.

## 4-11 ارزیابی

در نهایت با استفاده از مدل‌های یادگیرنده تقویتی بیان شده و اجرای شبیه‌سازی بر روی آن‌ها برای یادگیری آن‌ها و سپس جمع زدن میزان پاداش‌ها می‌توان مدل‌ها را تست نمود. سه عامل مورد تست قرار گرفته‌اند که به صورت زیر نتایج تست آن‌ها آمده است.

جدول 4- امتیازهای عامل‌های مختلف

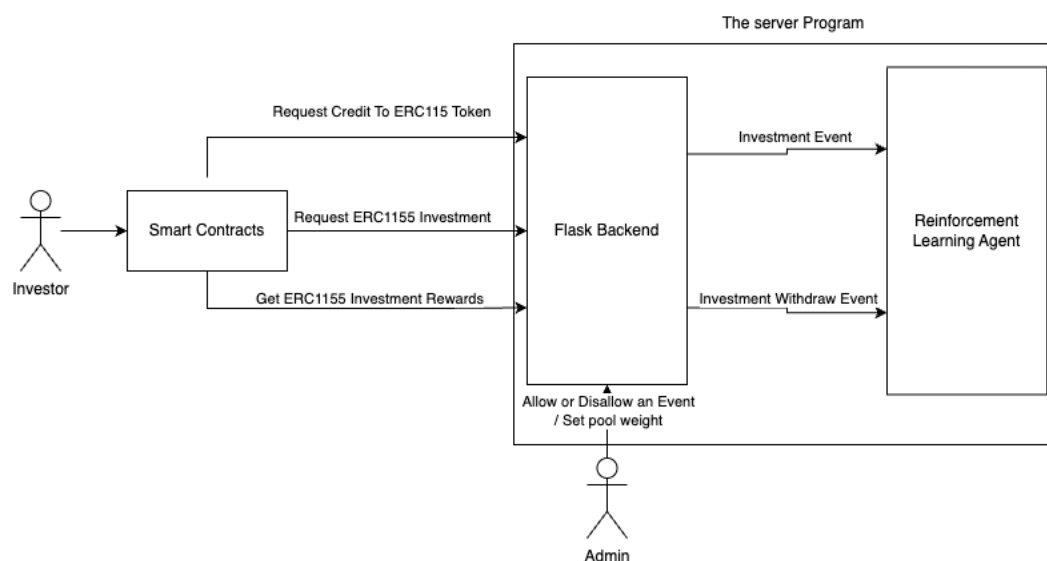
عامل رندوم	عامل DQN	عامل Sarsa
10981.15778705092	28328.062903721406	25753.4734386721

میزان زمان مورد نیاز برای یادگیری عامل بر حسب ثانیه به صورت زیر است:

عامل Sarsa	عامل DQN
453.0973138809204	48.10773420333862

## 4-12 نحوه پیاده‌سازی

معماری نهایی از حیث پیاده‌سازی به صورت زیر است:



تصویر 6 - معماری پیاده‌سازی

مراحل به صورت زیر است:

1. ابتدا سرمایه‌گذار با استفاده از قراردادهای هوشمند درخواست تبدیل اعتبار خود به ERC1155 را صادر می‌کند.
2. این درخواست به سرور Backend می‌رسد. در صورتی که کاربر مدیر در سیستم عمل تبدیل را طبق سیاست‌های خود تایید نموده باشد تبدیل اعتبار به توکن ERC1155 صورت می‌گیرد.
3. کاربر با توکن ERC1155 خود اعمال مختلفی از جمله سرمایه‌گذاری، سوزاندن، فروختن می‌تواند انجام دهد.



4. سرمایه‌گذار با استفاده از توکن خود می‌تواند در سرور درخواست سرمایه‌گذاری به سرور Flask بدهد.

5. در صورت تایید و تعیین وزن‌های استخر مربوطه توکن اعتبار سرمایه‌گذار در استخر قرار می‌گیرد و سرور Flask با ارتباط با عامل یادگیری تقویتی خبر سرمایه‌گذاری را به آن می‌دهد.

6. سرمایه‌گذار هرگاه که بخواهد سود خود را بردارد این امکان را دارد. سرور Flask با توجه به میزان سود که عامل یادگیرنده تقویتی تعیین می‌کند، میزان سود را برمی‌گرداند.

هر استخری سود سرمایه‌گذاری اعتبار رمز ارزی خاص خود را به عنوان خروجی می‌دهد و هر یک این اعتبار ها نیز ارزش ذاتی خاص خود را دارند. برای اینکه این عامل را در پیاده‌سازی دخیل کنیم راهکار زیر را ارائه کردیم.

به هر اعتبار کربن بنا به پارامترهای خود یک عدد اعتبار نسبت می‌دهیم. به هر یک از موارد استاندارد، نوع کربن، سال تولید یک وزن نسبت می‌دهیم که میزان سهم آن پارامتر در اعتبار را نشان می‌دهد. سپس برای برای هر مقدار پارامترها یک عددی را به عنوان میزان ارزش آن تعیین می‌کنیم. این وزن‌ها و نوع تعریف آن‌ها به صورت زیر است.

لازم به ذکر است که واحد زمانی‌ای برای انجام اعمالی مانند چک کردن قیمت و یادگیری عامل لازم است که در این پروژه سه ساعت در نظر گرفته شده است. بنا به نیاز می‌تواند این سه ساعت متغیر شود.

جدول 6- جدول امتیاز دهی به استخرهای اعتبارات کربن رمز ارزی

سال تولید	نوع کربن	استاندارد	
سال منهای 1985 ضرب در ۱۰۰	نگاشت از نام به عددی طبیعی	نگاشت از نام آن به عددی طبیعی	نحوه تعیین اعتبار
۰.۲	۰.۳	۰.۵	وزن‌ها

در نهایت تمام این اعتبارات در وزن‌هایشان ضرب می‌شوند و میزان اعتبار نوع اعتبار کربن را تعیین می‌کنند. یعنی:

$$\text{rep} = (\text{year} - 1985) * 0.2 + (\text{carbon type reputation}) * 0.3 + (\text{standard reputation}) * 0.5$$

در نهایت میزان فرکانس تولید شده توسط عامل یادگیرنده بر عدد اعتبار تقسیم می‌شود تا میزان اعتبار بر نرخ تولید تاثیر گذار باشد.

چند نمونه از جدول نگاشت برای استاندارد و نوع کربن در جدول‌های زیر آمده است.

جدول 7 - جدول نگاشت استاندارد به اعتبار

اعتبار	استاندارد
2000	Verra
1000	ساتبا

جدول 8 - جدول نگاشت نوع کربن به اعتبار

اعتبار	نوع کربن
۱۰۰۰	انرژی تجدیدپذیر
1200	جنگل‌زایی

#### 4-12-1 پیاده‌سازی قراردادهای هوشمند

چالش اصلی پیاده‌سازی قراردادهای هوشمند در مصرف کم حافظه و همچنین انجام تماس با بیرون است. برای چالش اول قواعد زیر برای ذخیره‌سازی داده‌ها ارائه شده است.

جدول 9 - جدول نگهداری اطلاعات رمزارزهای ERC1155

سال صدور اعتبار	میزان جبران کربن	نوع کربن	استاندارد
Year - 1985	Second and third 2 bytes	Second 4 bits	First 4 bits

با اینکار همه داده‌های مورد نیاز در یک متغیر ۳۲ بیتی جا می‌شوند.

برای تماس با بیرون از تماس oracle و محیط تست kovan استفاده شده است.

#### 2-12-4 پیاده‌سازی سرور backend

این سرور با استفاده از فناوری Flask طراحی شده تا به نیازهای قراردادهای هوشمند برای اخذ داده و همچنین ارتباط آن‌ها با یادگیرنده تقویتی تعبیه شده است.

#### 4-12-3 پیاده‌سازی یادگیری تقویتی

در هر یک از مدل‌ها از فناوری PyTorch و numpy و مجموعه Deque از کتابخانه پایتون استفاده شده است.

### 4-13 ارزیابی‌ای پس از اعمال تفاوت‌های استخرها

همان‌طور که گفته شد استخرهای سرمایه‌گذاری رمزارزی بنا به نوع ارزی که در خود نگهداری می‌کنند متفاوت هستند. همچنین افراد مختلف می‌توانند با الگوهای مختلفی سرمایه‌گذاری انجام دهند. پس دو سناریو تست برای مشاهده نحوه عملکرد عامل ارائه می‌کنیم و نتیجه را می‌بینیم.

#### 14-4 سرمایه‌گذاری دو نفر با الگوهایی یکسان در استخرهای متفاوت

##### 1-14-4 سناریوی اول

فرض کنیم در دو استخر با مشخصات Verra برای استاندارد، انرژی تجدید پذیر برای نوع کربن، سال ارائه اعتبار 2020 و ساتبا برای استاندارد، انرژی تجدید پذیر برای نوع کربن، سال ارائه اعتبار 2020 داریم. در اولی فردی با الگوی افزایش خطی و که از 100 تا 1100 می‌رود، سرمایه‌گذاری می‌کند. امتیاز استخر اول برابر زیر است:

$$\text{Pool1 score} = \text{verra score} * 0.5 + \text{renewable energy score} * 0.3 + (\text{year} - 2020) * 100 * 0.2 = 2000$$

$$\text{Pool2 score} = \text{satba score} * 0.5 + \text{renewable energy score} * 0.3 + (\text{year} - 2020) * 100 * 0.2 = 1500$$

نسبت این دو ۴ به ۳ است می‌شود. پس انتظار می‌رود در صورتی که هردوی این‌ها سرمایه‌گذاری کنند میزان ارزش در دست شخص دوم به شخص اول، ۴ به ۳ باشد. در نتیجه انجام شبیه‌سازی 208372 توکن برای شخص اول و 288341 توکن برای شخص دوم تولید شده است. که نسبت بین توکن شخص دوم به اول حدود 1.38 است که نسبت مناسبی است.

## 4-14-2 سناریوی دوم

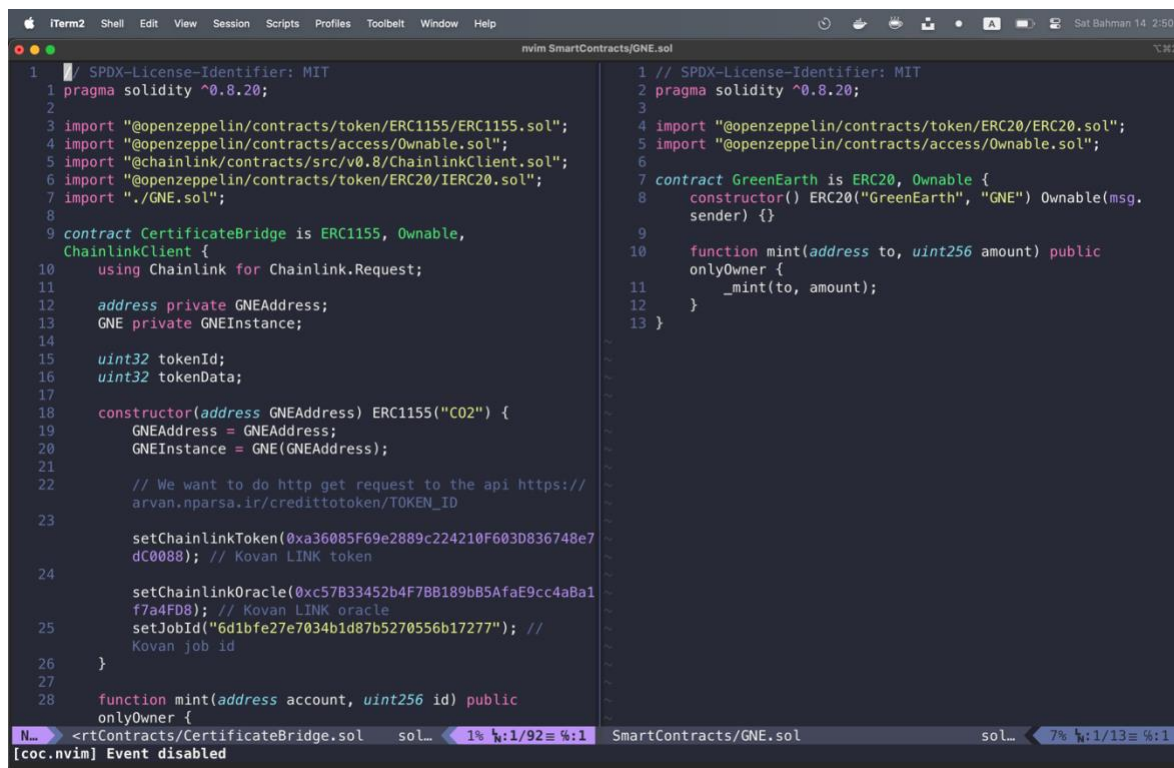
اگر چهار نفر باشند که هر یک در هر واحد زمانی با احتمالی از توزیع نرمال با میانگین 30 و انحراف معیار 10 سرمایه‌گذاری کنند و قیمت نیز روی ۱۰۰ ثابت باشد و استخر با مشخصات Verra برای استاندارد، انرژی تجدید پذیر برای نوع کربن، سال ارائه اعتبار 2020 باشد، و ۱۰۰۰ واحد زمانی اجرا انجام شود، میزان توکنی که دست هر یک از آنان در نهایت است چقدر می‌شود؟ در نهایت انتظار می‌رود که هر شخص حدود  $10 * 1000$  توکن سرمایه‌گذاری کرده باشد و انتظار می‌رود که هر شخص حدود ربع کل سرمایه‌گذاری را برده باشد. پس از اجرای شبیه‌سازی می‌بینیم که حدود 44445۱ توکن تولید شده و شخص اول 133112 و شخص دوم 91823 و شخص سوم 125,712 و شخص چهارم 93804 توکن گرفته که به نسب معقول برای این سناریو اعداد معقولی هستند.

## 4-15 جمع‌بندی

می‌توان با استفاده از مدل یادگیری تقویتی به حل مسئله استخر سرمایه‌گذاری رمزارزی پرداخت. شبکه‌های یادگیرنده تقویتی عمیق Q می‌توانند گزینه مناسبی برای حل این دسته مسائل که پیچیدگی‌های فراوانی دارند مورد استفاده قرار بگیرند. همچنین استفاده از الگوریتم پیشنهادی هیبریدی منجر به عملکرد بهینه‌تر یادگیرنده می‌شود. برای ارزیابی این مسئله نیاز به یک محیط پویا و تعامل‌دار با یادگیرنده بود. که این محیط ساخته شد. در نهایت با استفاده از تماس‌های Oracle و سرور Flask قراردادهای هوشمند به یادگیرنده و سرویس‌های لازم متصل شدند.

## پیوست

تصویر کدهای قراردادهای هوشمند:



The image shows a code editor with two files open. The left file is `CertificateBridge.sol` and the right file is `GreenEarth.sol`. Both files are Solidity contracts.

```
1 // SPDX-License-Identifier: MIT
1 pragma solidity ^0.8.20;
2
3 import "@openzeppelin/contracts/token/ERC1155/ERC1155.sol";
4 import "@openzeppelin/contracts/access/Ownable.sol";
5 import "@chainlink/contracts/src/v0.8/ChainlinkClient.sol";
6 import "@openzeppelin/contracts/token/ERC20/IERC20.sol";
7 import "../GNE.sol";
8
9 contract CertificateBridge is ERC1155, Ownable,
ChainlinkClient {
10     using Chainlink for Chainlink.Request;
11
12     address private GNEAddress;
13     GNE private GNEInstance;
14
15     uint32 tokenId;
16     uint32 tokenIdData;
17
18     constructor(address GNEAddress) ERC1155("C02") {
19         GNEAddress = GNEAddress;
20         GNEInstance = GNE(GNEAddress);
21
22         // We want to do http get request to the api https://
arvan.nparsa.ir/credittotoken/TOKEN_ID
23
24         setChainlinkToken(0xa36085F69e2889c224210F603D836748e7
dC0088); // Kovan LINK token
25
26         setChainlinkOracle(0xc57B33452b4F7BB189bB5AfaE9cc4aBa1
f7a4FD8); // Kovan LINK oracle
27         setJobId("6d1bfe27e7034b1d87b5270556b17277"); //
Kovan job id
28     }
29
30     function mint(address account, uint256 id) public
onlyOwner {
31
32     }
33 }
```

```
1 // SPDX-License-Identifier: MIT
2 pragma solidity ^0.8.20;
3
4 import "@openzeppelin/contracts/token/ERC20/ERC20.sol";
5 import "@openzeppelin/contracts/access/Ownable.sol";
6
7 contract GreenEarth is ERC20, Ownable {
8     constructor() ERC20("GreenEarth", "GNE") Ownable(msg.
sender) {}
9
10     function mint(address to, uint256 amount) public
onlyOwner {
11         _mint(to, amount);
12     }
13 }
```

بخشی از کدهای عامل یادگیرنده عمیق:

```

class DQNAgent:
    def __init__(self, input_size, output_value, learning_rate=0.001, epsilon=1.0, epsilon_decay=0.9995,
                 epsilon_min=0.01, memory_size=1000000, batch_size=64):
        self.input_size = input_size
        self.output_size = output_value * 2 + 1
        self.output_value = output_value
        self.learning_rate = learning_rate
        self.epsilon = epsilon
        self.epsilon_decay = epsilon_decay
        self.epsilon_min = epsilon_min
        self.memory_size = memory_size
        self.batch_size = batch_size
        self.memory = deque(maxlen=self.memory_size)
        self.model = DQN(self.input_size, self.output_size)
        self.optimizer = optim.Adam(self.model.parameters(), lr=self.learning_rate)
        self.loss = nn.MSELoss()
        self.last_reward = None
        self.goods = 0
        self.bads = 0

    def act(self, state):
        if self.epsilon <= self.epsilon_min:
            theta = np.random.beta(self.goods + 1, self.bads + 1)
            if 0.1 > theta:
                with torch.no_grad():
                    state = torch.tensor(state, dtype=torch.float).view(*shape, -1, self.input_size)

```

لینک کدهای زده شده:

<https://github.com/parsanoori/RLEnergy>

## منابع

- [1] United Nations "United Nations Treaties," [متصل]. 2015 Available: [https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg\\_no=XXVII-7-d&chapter=&27clang=en](https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=&27clang=en).
- [2] M. G. Kelley Hamrick, "Forest Trends," [متصل]. 2017 Available: [https://www.forest-trends.org/wp-content/uploads/07/2017/doc\\_5591.pdf](https://www.forest-trends.org/wp-content/uploads/07/2017/doc_5591.pdf).
- [3] Popular Science. [متصل]. 2022, Available: <https://www.popsci.com/environment/crypto-carbon-credit-tokens/>.
- [4] A. R. CHOW, "Time," [متصل]. 2022 Available: <https://time.com/6181907/crypto-carbon-credits/>.
- [5] Toucan, "Toucan | Carbon Market Infrastructure for climate action," [متصل]. Available: <https://toucan.earth/>.
- [6] Energy Web Origin, "EW-Origin," [متصل]. Available: <https://energy-web-foundation-origin.readthedocs-hosted.com/en/latest/>.
- [7] R. Sutton و A. Barto, Reinforcement Learning: An Introduction, MIT Press. 1998.

- [8] [K. K. D. S. A. G. I. A. D. W. M. R. Volodymyr Mnih , "Playing Atari with Deep Reinforcement Learning," Arxiv.2013 ,](#)
- [9] [A. G. . Richard S. Sutton , "Reinforcement Learning: An Introduction , Bradford Books , .1992](#)