Predict Chain

Matthew Pisano, Connor Patterson, William Hawkins

Spring 2023

1 Problem Summary

One of the main issues with artificial intelligence training and development today is accessibility. Oftentimes, individuals or groups possess data that they would wish to be used in predictive analysis. However, these people may not have access to the compute capacity to train predictive models on this data. Additionally,yet other people have neither access to predictive training data, nor do they have access to adequate computational resources. Compounding this issue is the fact that the entities that *do* have these computational resources often keep the results of trained models hidden from the public. Due to this lack of accessibility, similar models are often trained on similar datasets, obtaining similar results. This results in both a notable inefficiency, and the withholding of useful predictions from the majority of people.

2 Proposed Solution

2.1 Description

To help to aid with this issue, we propose PredictChain, a blockchain-based marketplace for predictive AI models. Through PredictChain, users are able to upload datasets for training predictive models, request that train models be trained on any previously uploaded datasets, or submit queries to those trained models. These various models will be operated by a central node or nodes with computing resources available. A variety of models will be made available, ranging from cheap, fast, and simple to more expensive, slower, and more powerful. This will allow for a large variety of predictive abilities for both simple and complex patterns. All the past predictions form these models will be stored on the blockchain for public viewing.

2.2 Implementation

The core of PredictChain is implemented using Python and its Algorand SDK. We have also developed a front end for the project in React. PredictChain is implemented using two primary parts: the client and the oracle. The client interacts with the end user through a front end website. It takes in user requests and reports back the results of those requests. The oracle handles the majority of the backend functionality. It handles the storing of uploaded datasets, the training of models on those datasets, the querying of those models, and the management of payments and rewards to the users.

In order to facilitate communication between the client and oracle, we use transactions on the Algorand blockchain. This is done through JSON-encoded nodes that are attached to the transactions. The advantage to using transactions as our primary mode of internode communication is that it creates an immutable, public record of all requests that have been made and all responses that have been returned. This ensures that, even though users may not have direct access to sensitive datasets they can still use the predictions that are produced from that dataset.

2.3 Addressing the Problem

Our project fulfils the need for a more accessible ML environment in two ways. It encourages users to generate high quality content through the usage of incentives and it ensures that the results of this effort are also publicly viewable. Using this system, people will be able to have access to high quality models to

make predictions based off of their datasets and people are able to view these predictions, cutting down on wasteful, duplicate work.

When users upload their datasets, they allow model to be trained on those datasets. Higher quality datasets will produce higher quality models. When users submit parameters for training, they allow the model that their parameters produce to be used publicly. Both of these users are rewarded for their work when a model is queried, and it produces a correct prediction. This encourages users to participate in contributing the resources needed for good predictions, while leaving a public record for other users to view through the usage of Algorand transactions.

3   Potential Impact

Given enough users, this project would have a notable impact on the state of machine learning usage. Currently, the problem of accessibility prevents many people from utilizing machine learning themselves, often leaving them to turn to highly private and centralized tech giants. This project would serve to open up this black-box of an industry and encourage the sharing of datasets and parameter configurations to create better models that are open to the public for usage.