A Scalable Visualization for Understanding Tree-based Ensemble Models

Tree-based ensemble models such as random forests and boosted trees are popular machine learning models. It is widely used in many application scenarios due to its practical effectiveness and is one of the most popular methods in data science competitions. It is composed of multiple decision trees. The result is determined by voting on the results of all decision trees.

Tree-based ensemble models have better performance than a single decision tree, but the interpretability is reduced the number of trees increases the branch structure become more complex the number of rules increases rapidly … Its lack of interpretability limits its use in high-stakes decisions, such as medical treatment, law enforcement, and financial forecasting. Visualization can usually provide good interpretability for models.

The project consists of three parts: backend, frontend, and model. The structure of the project is as follows:

```
backend --- a flask server
   app.py
frontend --- a visualization interface based on vue and vuetify
   src
   src/assets
   src/components --- vue components
   src/libs
   src/plugins
   src/store
   App.vue
   main.js
model
   data --- original dataset
   exp --- experiment code
   output --- hierarchical structure of rules
```

Set up the environment

The project needs python version >= 3.7 and node.js >= 14.0.

```
sudo apt install python3.7 python3.7-dev
curl -fsSL https://deb.nodesource.com/setup_16.x | sudo -E bash -
sudo apt-get install -y nodejs

pip install -r backend/requirements.txt
cd frontend
npm install
```

Run the backend

```
cd backend
export FLASK_APP=app.py
flask run --port=5000 --host=0.0.0.0
```

Run the frontend

```
cd frontend
npm run serve
```