

Гайд для проги Machine Learning

Назначение:

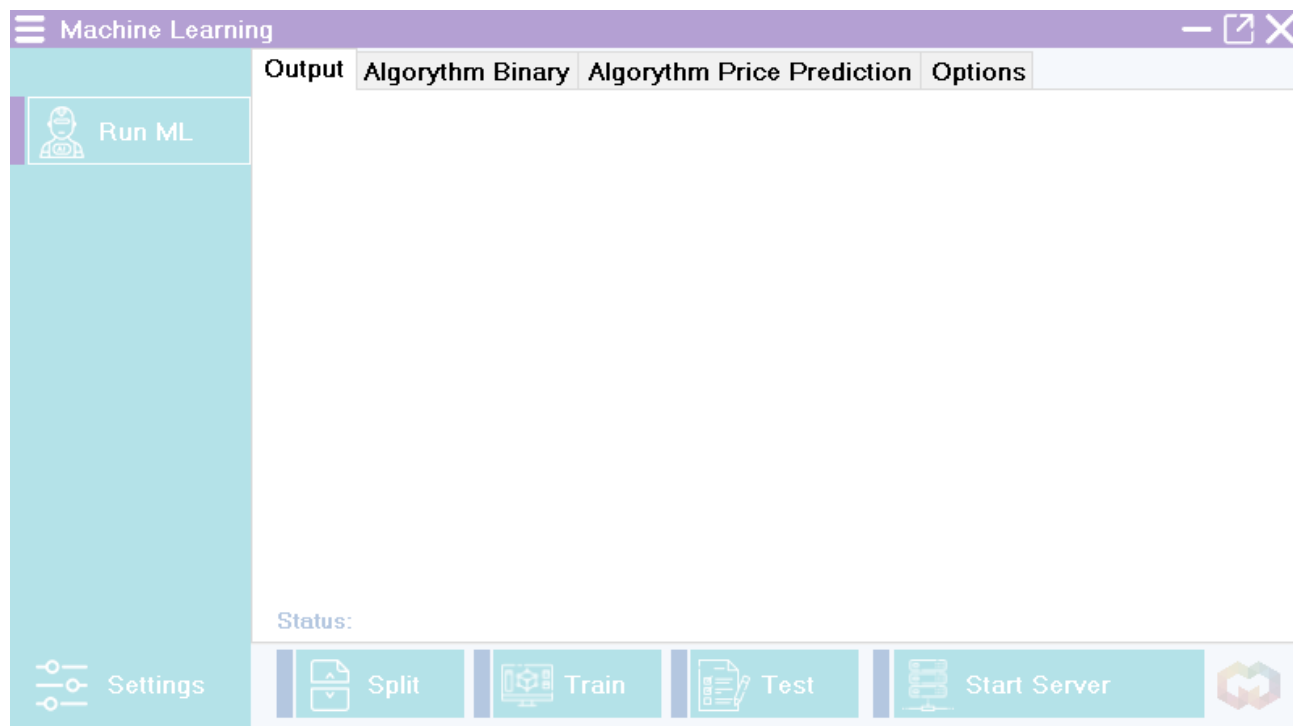
Прога решает две различные задачи машинного обучения:

- Бинарная классификация дает ответ true или false на основе входных данных
- Прогнозирование числового значения дает ответ в виде числового значения с плавающей точкой на основе входных данных

Формирование тренировочных данных:

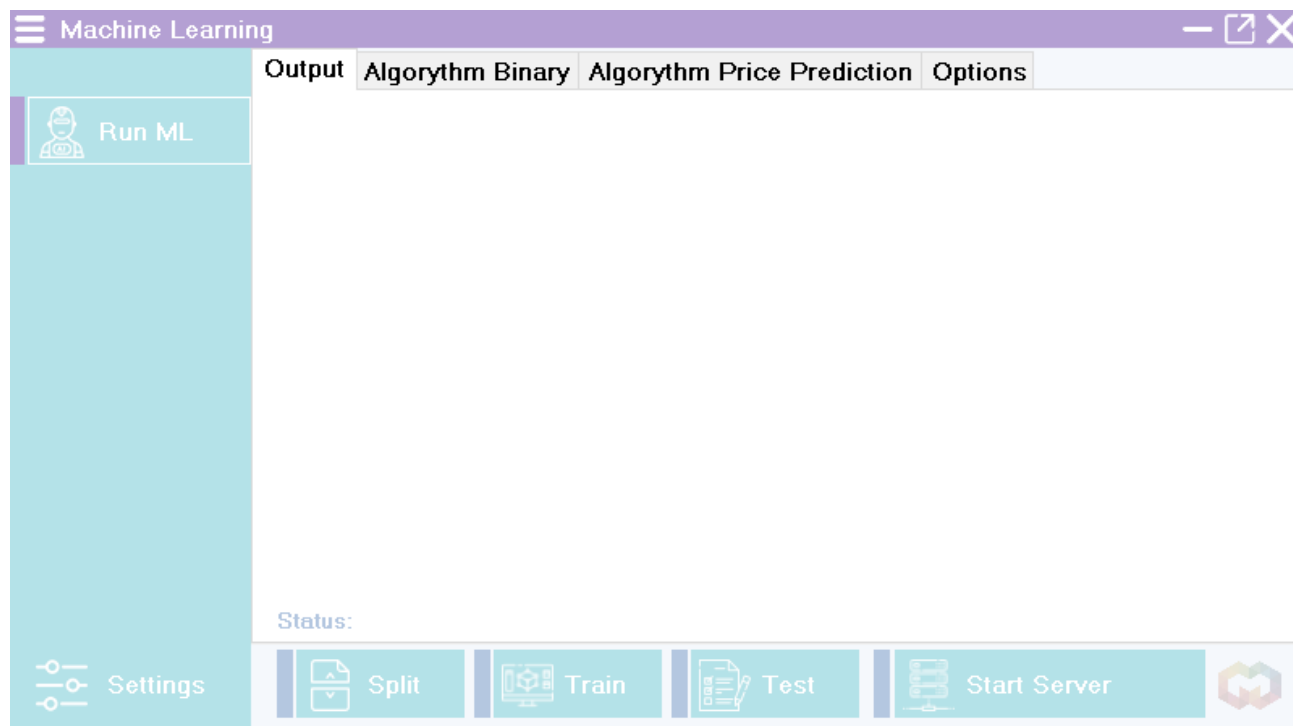
Программа для обучения работает с csv файлами в которых разделитель это запятая а делитель дробной части это точка. Файл должен иметь определенное название а именно {название валютной пары}-data-{шаг и вид обучения}.csv примеры BTCUSD-data-firstbin.csv BTCUSD-data-secondbin.csv BTCUSD-data-firstreg.csv BTCUSD-data-secondreg.csv. Файл должен иметь внутренний хэдер который является названием для каждой колонки входных данных. Первая колонка должна хранить результат для данного набора данных и называться Result все остальные колонки должны называться Val_1, Val_2...Val_99 в них хранятся числовые значения ваших индикаторов или чего угодно применимых для значения в колонке Result, всего количество колонок может быть от 2х (Result,Val_1) до 100 (Result,Val_1..Val_99) в зависимости от объема подаваемого набора данных. Кодировка файла особого значения не имеет так как эта версия проги не работает с анализом настроения текстовых цитат, или с адресами расположения графической или иной информации, а принимает лишь цифровые значения, но все же рекомендую использовать utf-8 или windows-1251.

Пример сформированного файла для прогноза числового значения:



где Result количество бар зигзага до следующего цифрового значения, а Val_1..Val_4 значения некоторых индикаторов в точке формирования узла зигзага.

Пример сформированного файла для прогноза бинарной классификации:



где Result булево значение являлся ли узел зигзага крайним перед перед обрантым узлом, а Val_1..Val_4 значения некоторых индикаторов в точке формирования узла зигзага.

Программа работает с 4мя наборами данных, два надора для бинарной классификации и два набора для прогнозирования числового значения. Пример применения на зигзаге:

BTCUSD-data-firstbin.csv — выборка для тренировки задачи определения является ли верхний узел зигзага крайним

BTCUSD-data-secondbin.csv - выборка для тренировки задачи определения является ли нижний узел зигзага крайним

BTCUSD-data-firstreg.csv - выборка для тренировки задачи прогнозирования конечной цены только что сформированного узла зигзага

BTCUSD-data-secondreg.csv - выборка для тренировки задачи прогнозирования количества бар только что сформированного узла зигзага

Программе начхать что именно тренировать важно лишь подводить свои задачи прогнозирования к бинарной классификации (да, нет) либо прогнозированию числового значения (какое то конкретное число например 123.00001), а общее количество задач не должно превышать четырех, два из которых бинарка и два число.

Отправка входных данных:

Данные должны быть отправлены на адрес и порт указанные в настройках программы и содержать в себе только значения индикаторов через запятую без пробелов, разделитель целой и дробной части точка, значение же Result программа вернет в качестве спрогнозированного ответа

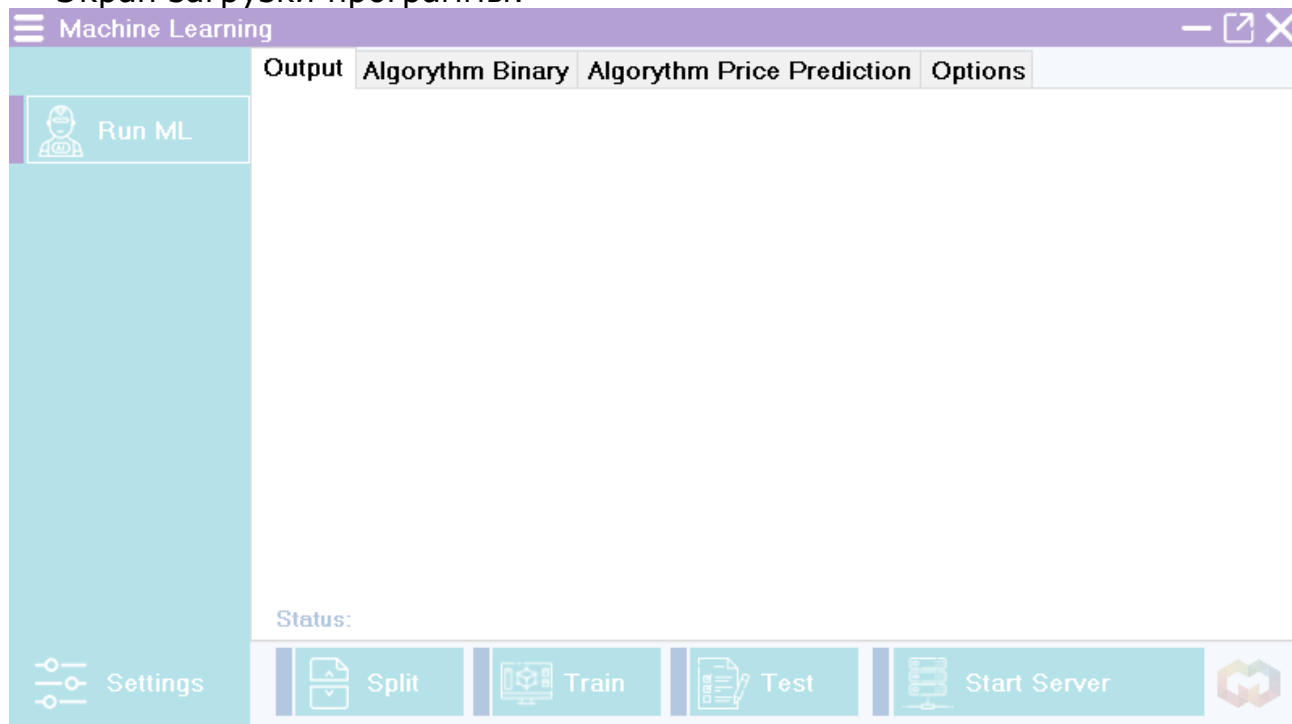
Получение выходных данных:

Советник который использует в своей торговле данные спрогнозируемые прогой получит ответ вида **False,True,1,1.13890** где первые два значения это ответы по первой и второй моделям бинарной классификации а третье и

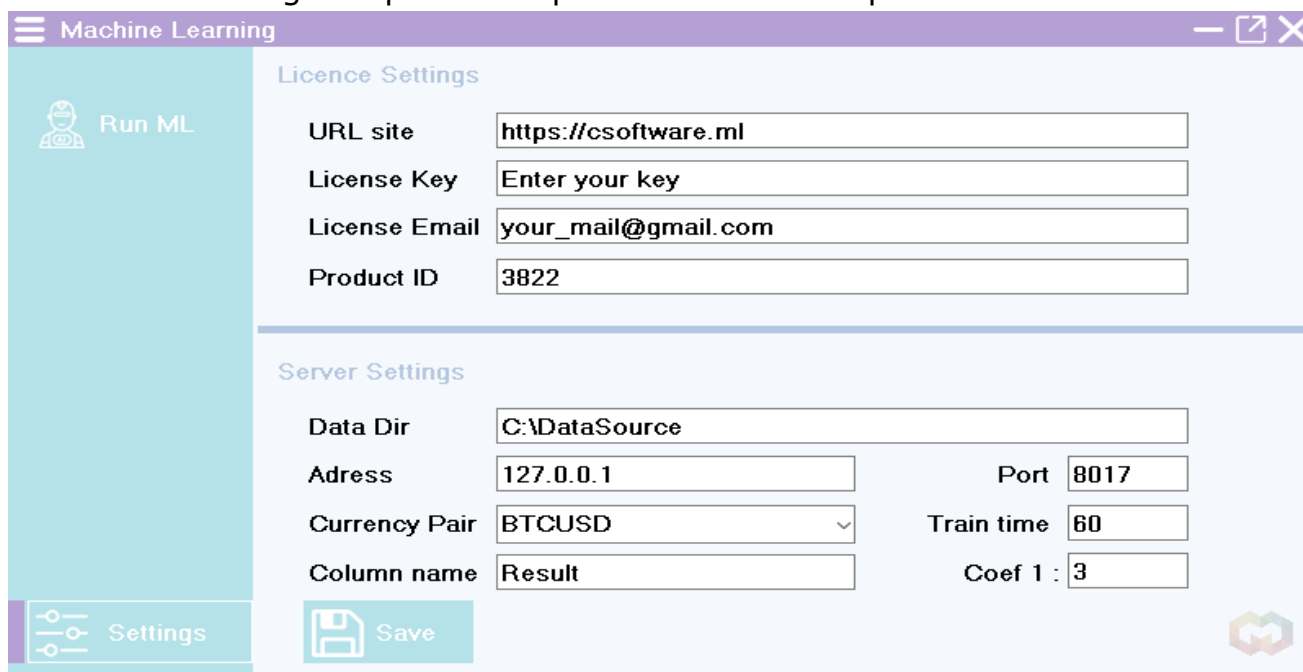
четвертое значение ответы по первой и второй моделям прогнозирования числового значения.

Графический интерфейс:

Экран загрузки программы:



Клавиша settings открывает экран основных настроек:



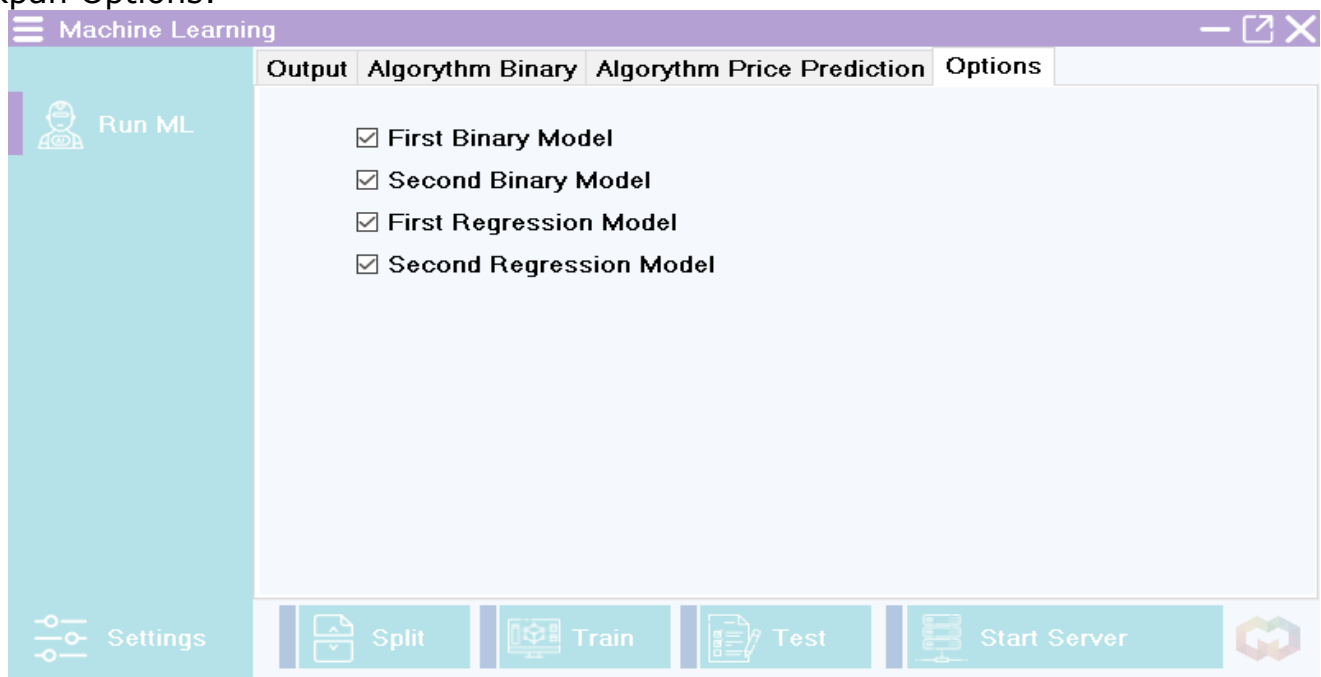
где:

- URL site – поле для ввода адреса сайта проверки лицензии
- License Key — ключ выданный сайтом после приобретения программы. Его можно найти в личном кабинете пользователя сайта в папке API Keys
- License Email – поле ввода почтового ящика на который зарегистрирован аккаунт купивший программу.
- Product ID – поле ввода ID программы. Его можно найти в личном кабинете

пользователя сайта в папке API Keys

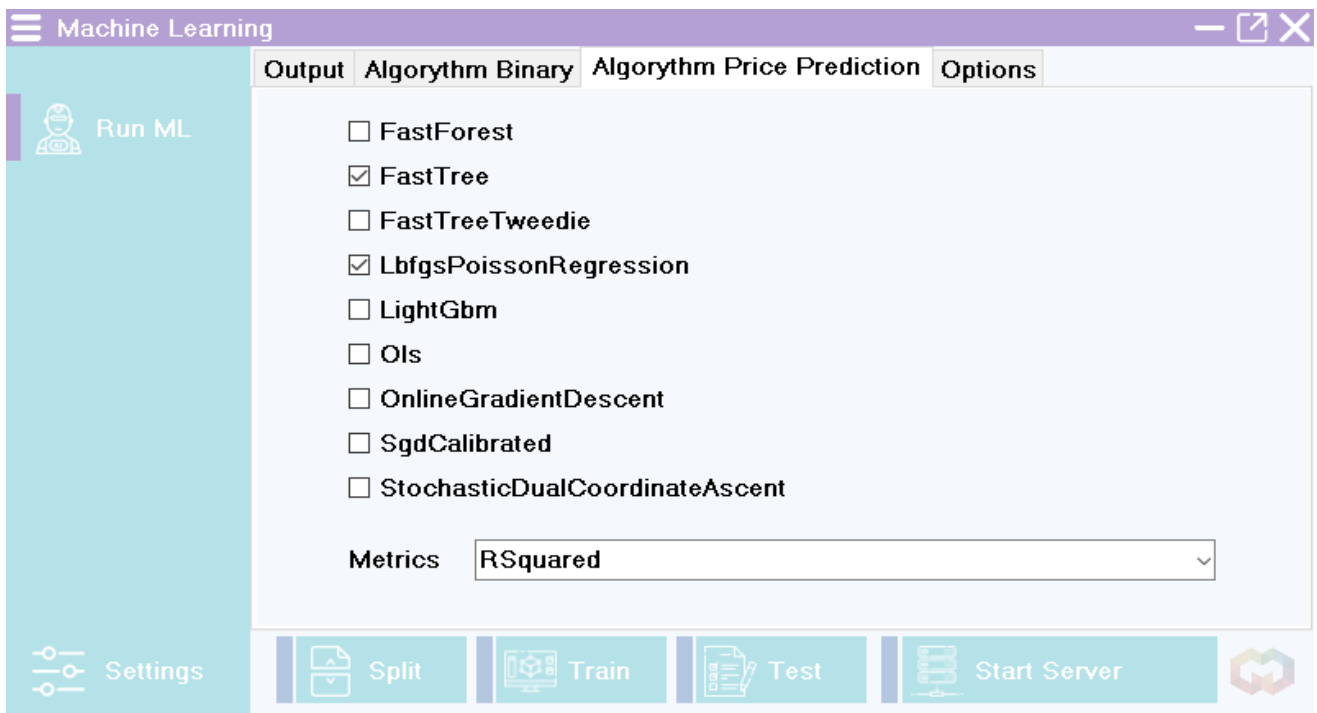
- Data Dir – поле ввода адреса папки с хранимыми предварительно спарсенными файлами выборками. Обратите внимание на то что программа не создает эти файлы или саму папку, эти данные генерируются советником под ваш терминал, программа лишь обрабатывает эти файлы и строит на их основе модель.
- Adress – поле ввода IP адреса на котором программа будет принимать входные данные для последующего анализа, и передачи ответа с спрогнозированным значением.
- Port - поле ввода номера свободного порта на котором программа будет принимать входные данные для последующего анализа, и передачи ответа с спрогнозированным значением.
- Currency name – поле выбора валютной пары с которой будет работать запущенный экземпляр программы.
- Column name – поле ввода названия столбца с калиброванными ответами. Оставляйте как есть не менять.
- Train time – время тренировки каждой из четырех моделей. Больше не значит лучше, время подбирается экспериментально.
- Coef 1: поле ввода коэффициента деления файлов с данными на тренировочную и тестовую выборки.
- Save – клавиша сохранения настроек.

Экран Options:



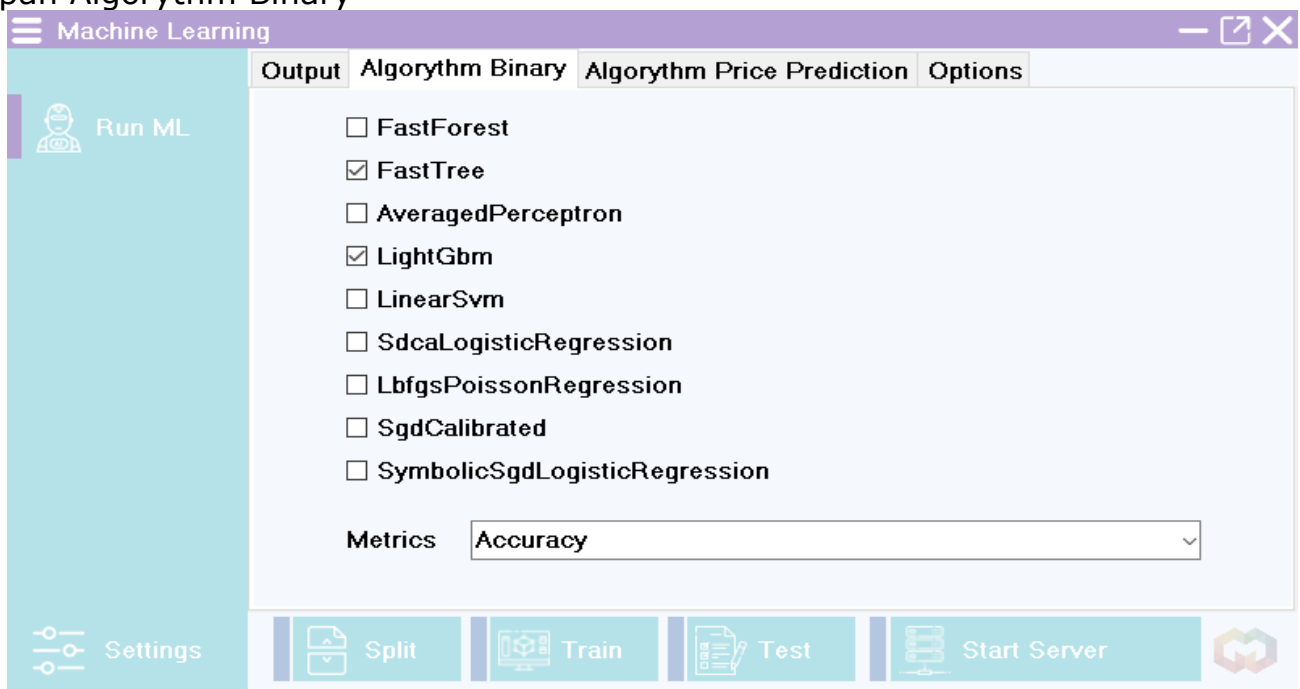
Экран включения/отключения тренировки моделей прогнозирования

Экран Algorhythm Price Prediction



Выбор алгоритмов прогнозирования а так же метрики относительно которой будет проходить оптимизация для выборок прогнозирования числового значения

Экран Algorithm Binary



Выбор алгоритмов прогнозирования а так же метрики относительно которой будет проходить оптимизация для выборок бинарной классификации

Экран Output

Machine Learning

Run ML

Settings

Output

Algoorythm Binary

Algoorythm Price Prediction

Options

The best results for the model: FastTreeBinary

Characteristics of the best model:

Accuracy: 0.750

F1Score (Usefull): 0.444

Negative Precision (How many selected items are relevant?): 0.765

Negative Recall (How many relevant items are selected?): 0.929

Positive Precision (How many selected items are relevant?): 0.667

Positive Recall (How many relevant items are selected?): 0.333

Characteristics of the best model on the test sample:

Accuracy: 0.868

F1Score (Usefull): 0.768

Negative Precision (How many selected items are relevant?): 0.953

Negative Recall (How many relevant items are selected?): 0.867

Positive Precision (How many selected items are relevant?): 0.686

Positive Recall (How many relevant items are selected?): 0.871

Status: Modeling completed.

Split

Train

Test

Start Server

Отображение основных метрик и данных моделирования.

Внимание! Для корректной работы программы необходимо поменять значение разделителя целой и дробной части в региональных настройках на точку.