


Руководство 2. Обучение моделей кредитных рисков в Студии машинного обучения (классической)

Статья • 11.02.2019

ОБЛАСТЬ ПРИМЕНЕНИЯ:  Студия машинного обучения (классическая) 

[Машинное обучение Azure](#)

Важно!

Поддержка Студии машинного обучения (классической) будет прекращена 31 августа 2024 г. До этой даты рекомендуется перейти на [Машинное обучение Azure](#).

Начиная с 1 декабря 2021 года вы не сможете создавать новые ресурсы Студии машинного обучения (классической). Существующие ресурсы Студии машинного обучения (классическая версия) можно будет использовать до 31 августа 2024 г.

- См. [сведения о переносе проектов машинного обучения из Студии машинного обучения \(классическая версия\) в Машинное обучение Azure](#).
- См. дополнительные сведения о [Машинном обучении Azure](#).

Прекращается поддержка документации по Студии машинного обучения (классическая версия). В будущем она может не обновляться.

В этом руководстве подробно описывается процесс разработки решения прогнозной аналитики. Мы создадим в Студии машинного обучения (классической) простую модель, а затем развернем ее в качестве веб-службы машинного обучения. Развернутая модель позволяет создавать прогнозы на основе новых данных. Это руководство представляет собой **вторую часть серии, состоящей из трех частей**.

Предположим, требуется спрогнозировать кредитный риск претендентов на кредит

на основе сведений, которые они предоставляют в заявке на кредит.

Оценка кредитных рисков — это сложная задача. Мы немного упростим ее для этого руководства. На примере этой задачи вы создадите решение прогнозной аналитики с помощью Студии машинного обучения (классической). Для этого вы используете Студию машинного обучения (классическую) и веб-службу машинного обучения.

В этом руководстве из трех частей описание начинается с общедоступных данных по кредитным рискам. Затем необходимо разработать и обучить модель прогнозирования. и, наконец, развернете модель в качестве веб-службы.

В [первой части учебника](#) вы создали рабочую область Студии машинного обучения (классической), отправили в нее данные и создали эксперимент.

Эта часть руководства состоит из таких этапов:

- ✓ обучение нескольких моделей;
- ✓ Оценка и анализ моделей

В [третьей части руководства](#) будет необходимо развернуть модель в качестве веб-службы.

Предварительные требования

Выполните действия, описанные в [первой части этого руководства](#).

обучение нескольких моделей;

Одно из преимуществ использования Студии машинного обучения (классической) для создания моделей машинного обучения заключается в том, что вы можете проверить несколько типов моделей и сравнить их результаты в рамках одного эксперимента. Этот тип экспериментов помогает найти наилучшее решение проблемы.

В эксперименте, разрабатываемом в этом руководстве, вы создадите два разных типа моделей и сравните результаты их оценки, чтобы решить, какой алгоритм использовать в финальном эксперименте.

Существует множество моделей, которые можно выбрать. Чтобы просмотреть доступные модели, разверните узел **Машинное обучение** в палитре модулей, а затем **Initialize Model** (Инициализация модели) и узлы под ним. Для этого эксперимента необходимо выбрать модули [Two-Class Support Vector Machine](#) (Двухклассовый метод опорных векторов) и [Two-Class Boosted Decision Tree](#) (Двухклассовое расширяемое дерево принятия решений).

В этот эксперимент мы добавим модули [Two-Class Boosted Decision Tree](#) (Двухклассовое расширяемое дерево принятия решений) и [Two-Class Support Vector Machine](#) (Двухклассовый метод опорных векторов).

Двухклассовое увеличивающееся дерево принятия решений;

Сначала настройте модель увеличивающегося дерева принятия решений.

1. Найдите модуль [Two-Class Boosted Decision Tree](#) (Двухклассовое расширяемое дерево принятия решений) на палитре модулей и перетащите его на холст.
2. Найдите модуль [Train Model](#) (Обучение модели), перетащите его на холст, а затем подключите точку вывода модуля [Two-Class Boosted Decision Tree](#) (Двухклассовое увеличивающееся дерево решений) к левому порту ввода модуля [Train Model](#) (Обучение модели).

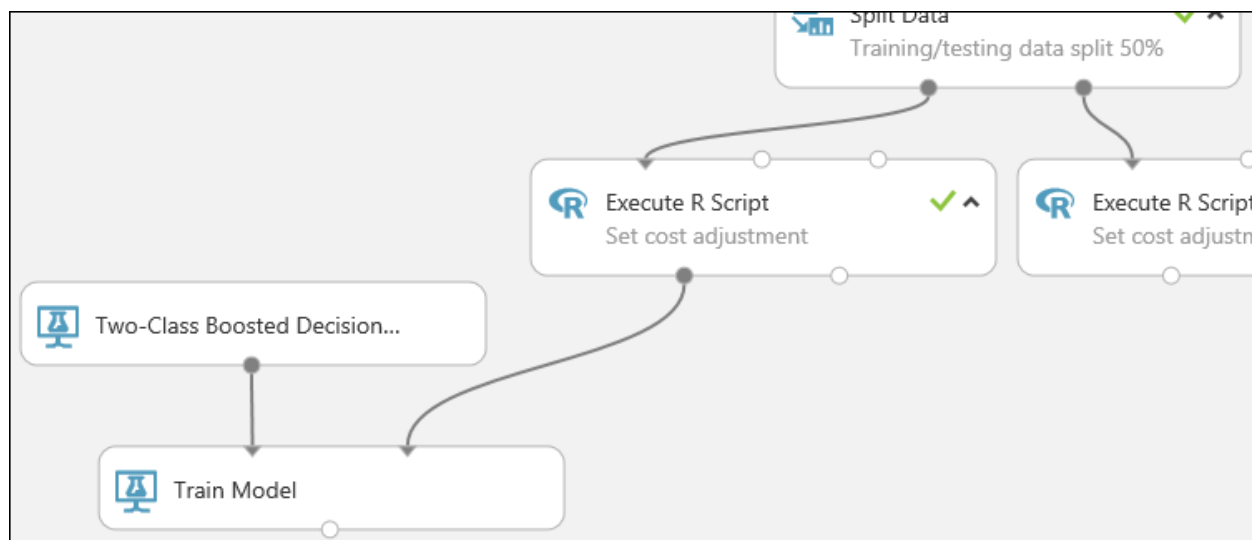
Модуль [Two-Class Boosted Decision Tree](#) (Двухклассовое расширяемое дерево принятия решений) инициализирует универсальную модель, а модуль [Train Model](#) (Обучение модели) использует обучающие данные для обучения модели.

3. Соедините левый выход левого модуля [Выполнить сценарий R](#) с правым портом ввода модуля [Обучение модели](#). Напомним, что в этом руководстве вы [использовали для обучения данные, поступающие от левой стороны](#) модуля [разделения данных](#).

Совет

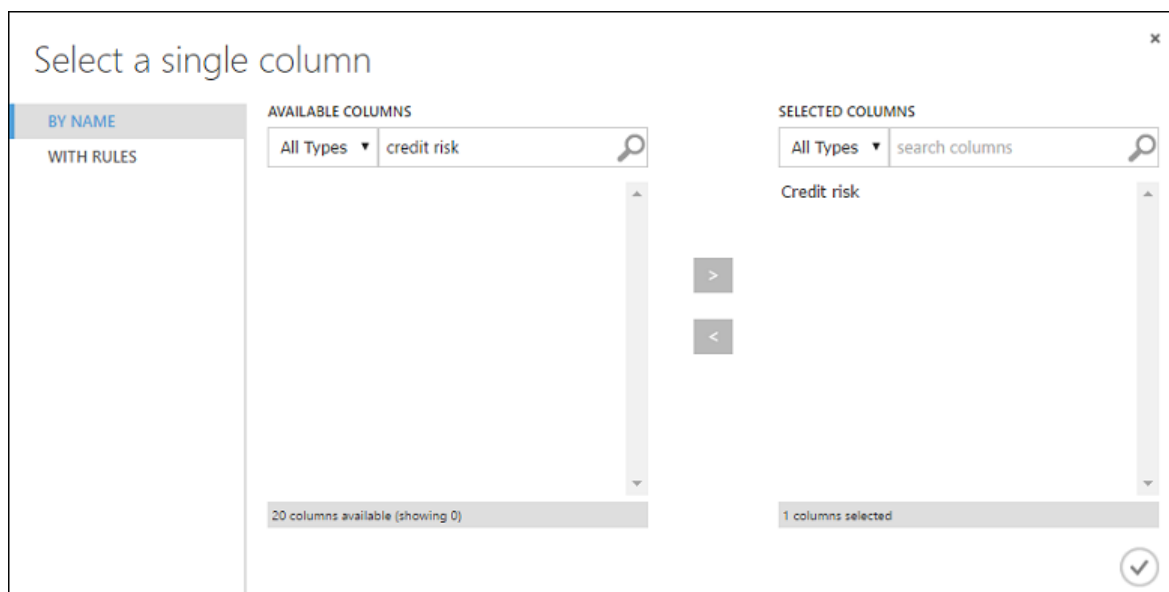
Для этого эксперимента вам не нужны два входа и один выход модуля [Выполнить сценарий R](#), поэтому их можно оставить неподключенными.

Теперь эта часть эксперимента выглядит следующим образом.



Теперь необходимо сообщить модулю **Train Model** (Обучение модели), что модель должна прогнозировать значения кредитного риска (поле "Credit Risk" (Кредитный риск)).

1. Выберите модуль **Train Model** (Обучение модели). В области **Properties** (Свойства) щелкните **Launch column selector** (Запустить средство выбора столбцов).
2. В диалоговом окне **Select a single column** (Выберите один столбец) введите в разделе **Available Columns** (Доступные столбцы) в поле поиска значение "Credit Risk", выберите внизу столбец Credit Risk (Кредитный риск) и нажмите кнопку со стрелкой вправо (>), чтобы переместить столбец "Credit Risk" (Кредитный риск) в раздел **Selected Columns** (Выбранные столбцы).



3. Нажмите кнопку с флажком **OK**.

Двухклассовая машина опорных векторов

Теперь настроим модель SVM.

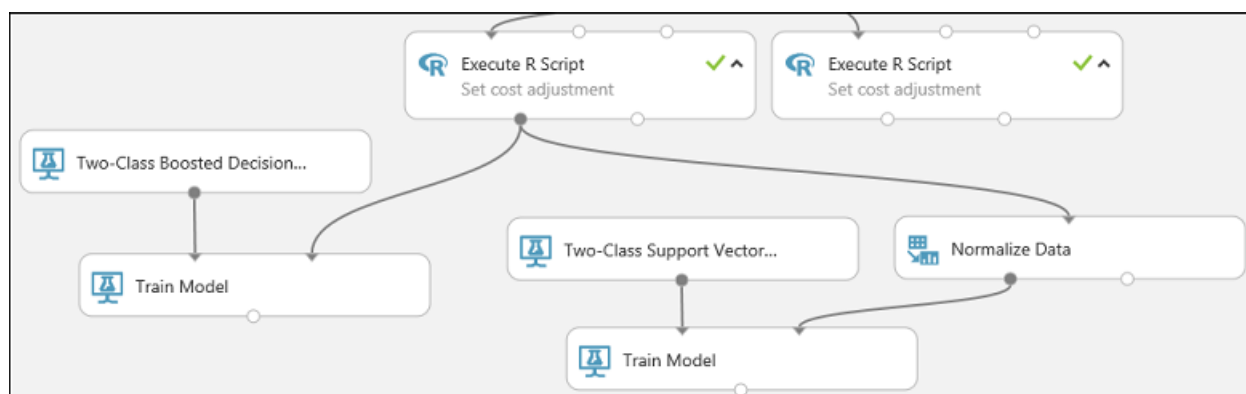
Сначала немного информации о SVM. "Повышенные" деревья принятия решения хорошо работают с атрибутами любого типа. Однако поскольку модуль SVM создает линейный классификатор, генерируемая им модель имеет наименьшую ошибку при проверке среди всех атрибутов одного масштаба. Чтобы свести все числовые признаки к одному масштабу, используйте преобразование Tanh (с помощью модуля [Normalize Data](#) (Нормализация данных)). Этот алгоритм преобразует все числа в значения, лежащие в интервале [0,1]. Модуль SVM преобразовывает строковые признаки в категориальные, а затем в двоичные признаки 0/1, поэтому вручную их преобразовывать не нужно. Кроме того, вам не нужно преобразовывать столбец кредитного риска (столбец 21) — он числовой, но мы обучаем модель прогнозировать именно его значение, поэтому его не нужно трогать.

Чтобы настроить модель SVM, выполните следующие действия:

1. Найдите модуль [Two-Class Support Vector Machine](#) (Двухклассовый метод опорных векторов) на палитре модулей и перетащите его на холст.
2. Щелкните правой кнопкой мыши модуль [Train Model](#) (Обучение модели), выберите **Копировать**, а затем щелкните холст правой кнопкой мыши и

- выберите **Вставить**. Обратите внимание, что копия модуля **Train Model** (Обучение модели) имеет тот же набор выбранных столбцов, что и оригинал.
3. Соедините вывод модуля **Two-Class Support Vector Machine** (Двухклассовый метод опорных векторов) с левым портом ввода второго модуля **Train Model** (Обучение модели).
 4. Найдите модуль **Normalize Data** (Нормализация данных) и перетащите его на холст.
 5. Соедините ввод этого модуля с левым портом вывода левого модуля **Execute R Script** (Выполнение скрипта R). Обратите внимание, что порт вывода любого модуля может быть подключен к нескольким модулям.
 6. Соедините левый порт вывода модуля **Normalize Data** (Нормализация данных) с правым портом ввода второго модуля **Train Model** (Обучение модели).

Теперь эта часть эксперимента должна выглядеть следующим образом.

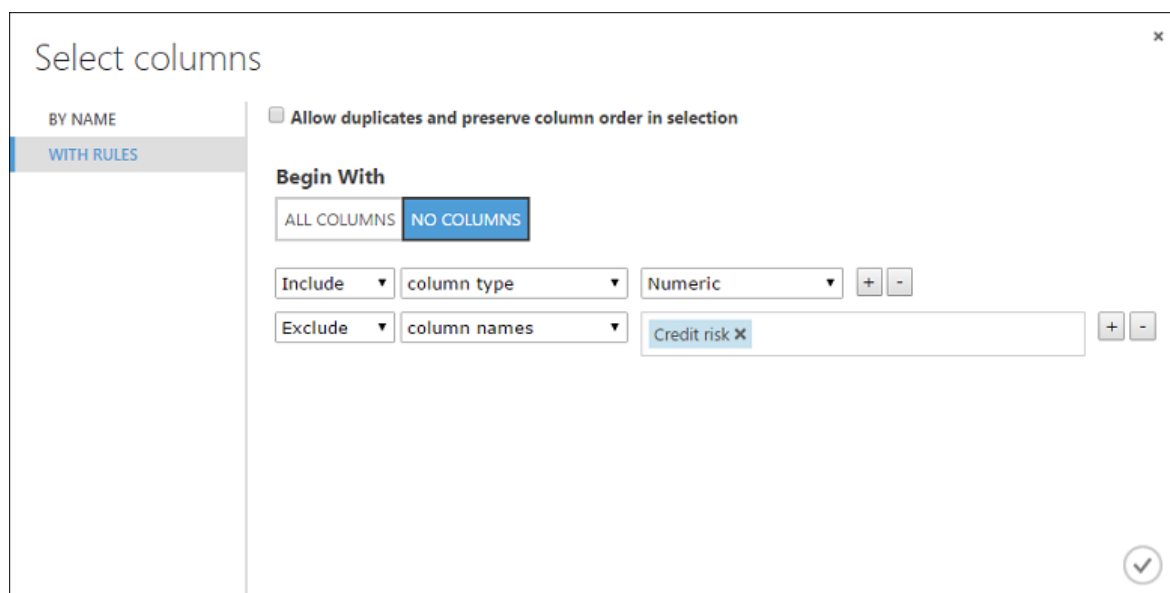


Теперь настройте модуль **Normalize Data** (Нормализация данных).

1. Щелкните модуль **Normalize Data** (Нормализация данных), чтобы выбрать его. На панели **Properties** (Свойства) выберите для параметра **Transformation method** (Метод преобразования) значение **Tanh**.
2. Щелкните **Запустить средство выбора столбцов** (Launch column selector), укажите для параметра **Begin With** (Начало с) значение **No columns** (Без столбцов). В первом раскрывающемся списке выберите пункт **Включение**, во втором — **column type** (тип столбца), а в третьем — **Числовой**. Это указывает, что будут преобразованы все числовые столбцы (и только числовые).
3. Щелкните знак плюса (+) в правой части этой строки. Это позволит создать

строку с раскрывающимися списками. Выберите в первом раскрывающемся списке значение **Exclude** (Исключить), во втором раскрывающемся списке выберите **column names** (Имена столбцов), затем введите текст "Credit risk" в текстовом поле. Так мы сообщим модулю, что столбец Credit Risk (Кредитный риск) нужно игнорировать. В противном случае этот столбец будет преобразован, так как является числовым.

4. Нажмите кнопку с флажком **OK**.



Теперь модуль [Normalize Data](#) (Нормализация данных) настроен для выполнения преобразования "Tanh" для всех числовых столбцов, за исключением столбца "Credit Risk" (Кредитный риск).

Оценка и анализ моделей

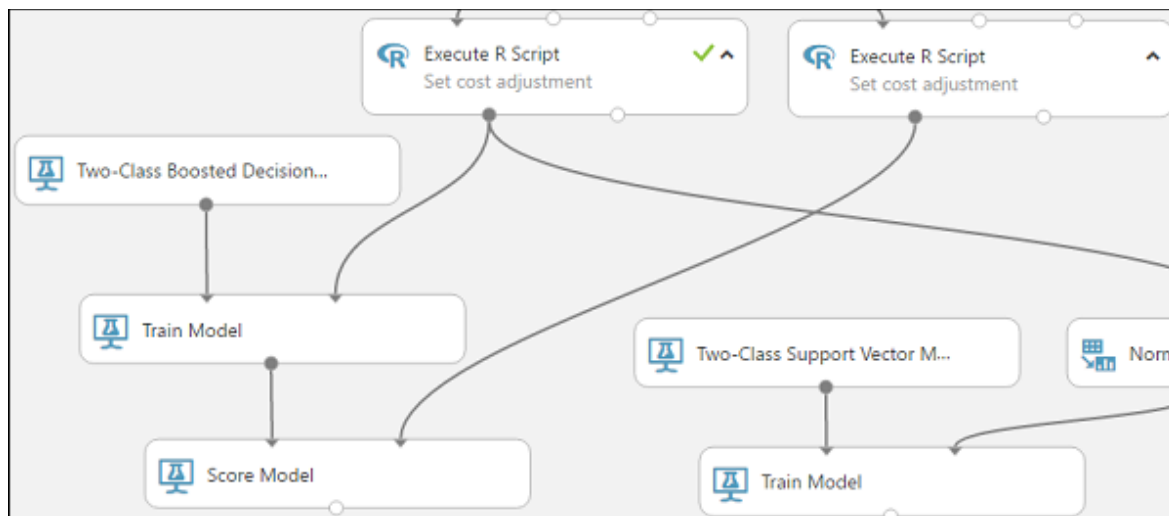
Вы будете использовать для оценки обученных моделей данные тестирования, которые были отделены модулем [Split Data](#) (Разделение данных). Затем можно сравнить результаты двух моделей, чтобы определить, какая из них создает лучший результат.

Добавление модулей "Score Model" (Оценка модели)

1. Найдите модуль [Score Model](#) (Оценка модели) и перетащите его на холст.
2. Подключите модуль [Train Model](#) (Обучение модели), который подключен к

модулю [Two-Class Support Vector Machine](#) (Двухклассовый метод опорных векторов), к левому порту ввода нового модуля [Score Model](#) (Оценка модели).

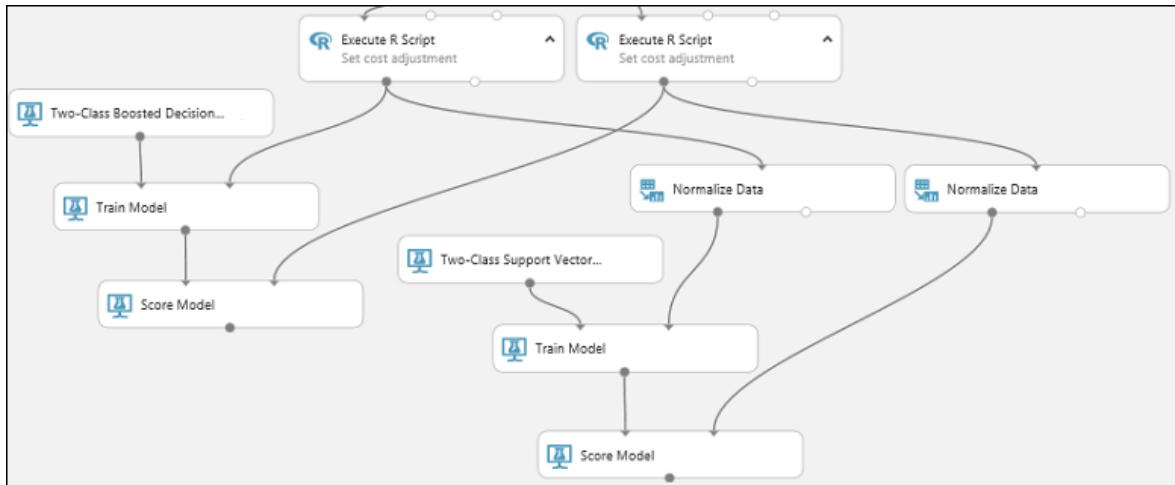
3. Соедините правый модуль [Execute R Script](#) (Выполнение сценария R), который возвращает данные для тестирования, с правым портом ввода модуля [Score Model](#) (Оценка модели).



Теперь модуль [Score Model](#) (Оценка модели) может получить сведения о кредите из тестовых данных, обработать их с помощью модели и сравнить созданные моделью прогнозы с фактическими данными столбца кредитного риска из тестовых данных.

4. Скопируйте и вставьте модуль [Score Model](#) (Оценка модели), чтобы создать вторую копию.
5. Соедините выход модуля SVM (т. е. порт вывода модуля [Train Model](#) (Обучение модели), подключенного к модулю [Two-Class Support Vector Machine](#) (Двухклассовый метод опорных векторов)) со входным портом второго модуля [Score Model](#) (Оценка модели).
6. Для модели SVM необходимо выполнить то же преобразование тестовых данных, что и для обучающих данных. Поэтому скопируйте и вставьте модуль [Normalize Data](#) (Нормализация данных), чтобы создать второй его экземпляр, и подключите его к выводу правого модуля [Execute R Script](#) (Выполнение сценария R).
7. Соедините левый вывод второго модуля [Normalize Data](#) (Нормализация данных) с правым портом ввода второго модуля [Score Model](#) (Оценка

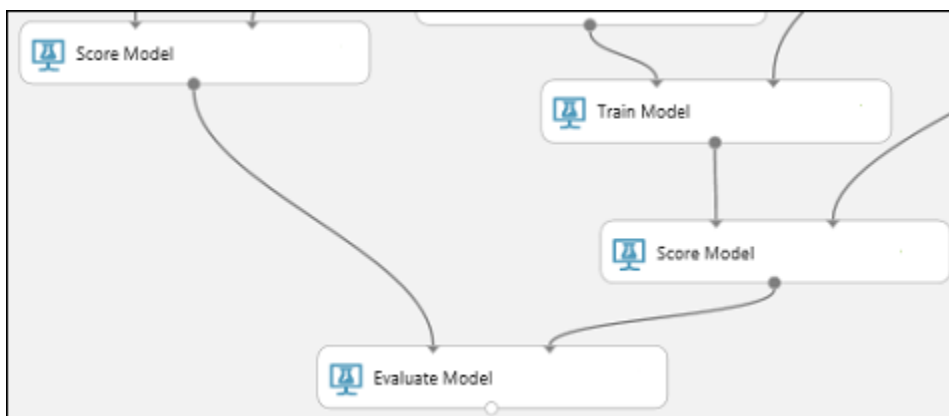
модели).



Добавление модуля "Evaluate Model" (Анализ модели)

Чтобы проанализировать и сравнить два результата оценки, используйте модуль [Evaluate Model](#) (Оценка модели).

1. Найдите модуль [Evaluate Model](#) (Оценка модели) и перетащите его на холст.
2. Подключите порт вывода модуля [Score Model](#) (Оценка модели), связанного с модулем увеличивающегося дерева решений, к левому входному порту модуля [Evaluate Model](#) (Анализ модели).
3. Соедините второй модуль [Score Model](#) (Оценка модели) с левым портом ввода.

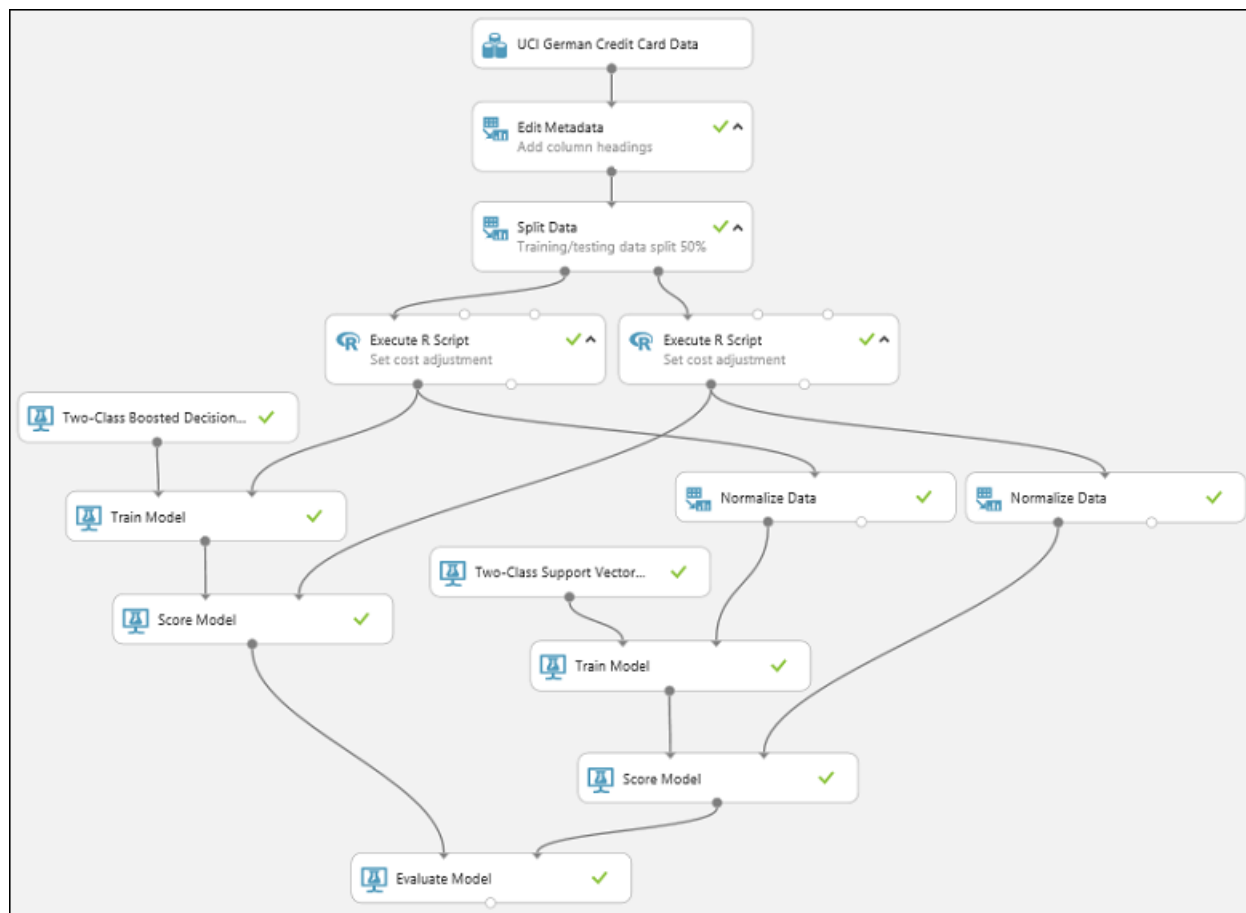


Запуск эксперимента и проверка результатов

Нажмите кнопку **ЗАПУСТИТЬ** внизу холста, чтобы запустить эксперимент. Это

может занять несколько минут. В каждом модуле будет отображаться вращающийся индикатор, показывающий, что модуль выполняется. Когда модуль завершит работу, появится зеленый флажок. Если эта галочка появилась во всех модулях, значит эксперимент завершил выполнение.

Теперь наш эксперимент должен выглядеть следующим образом:

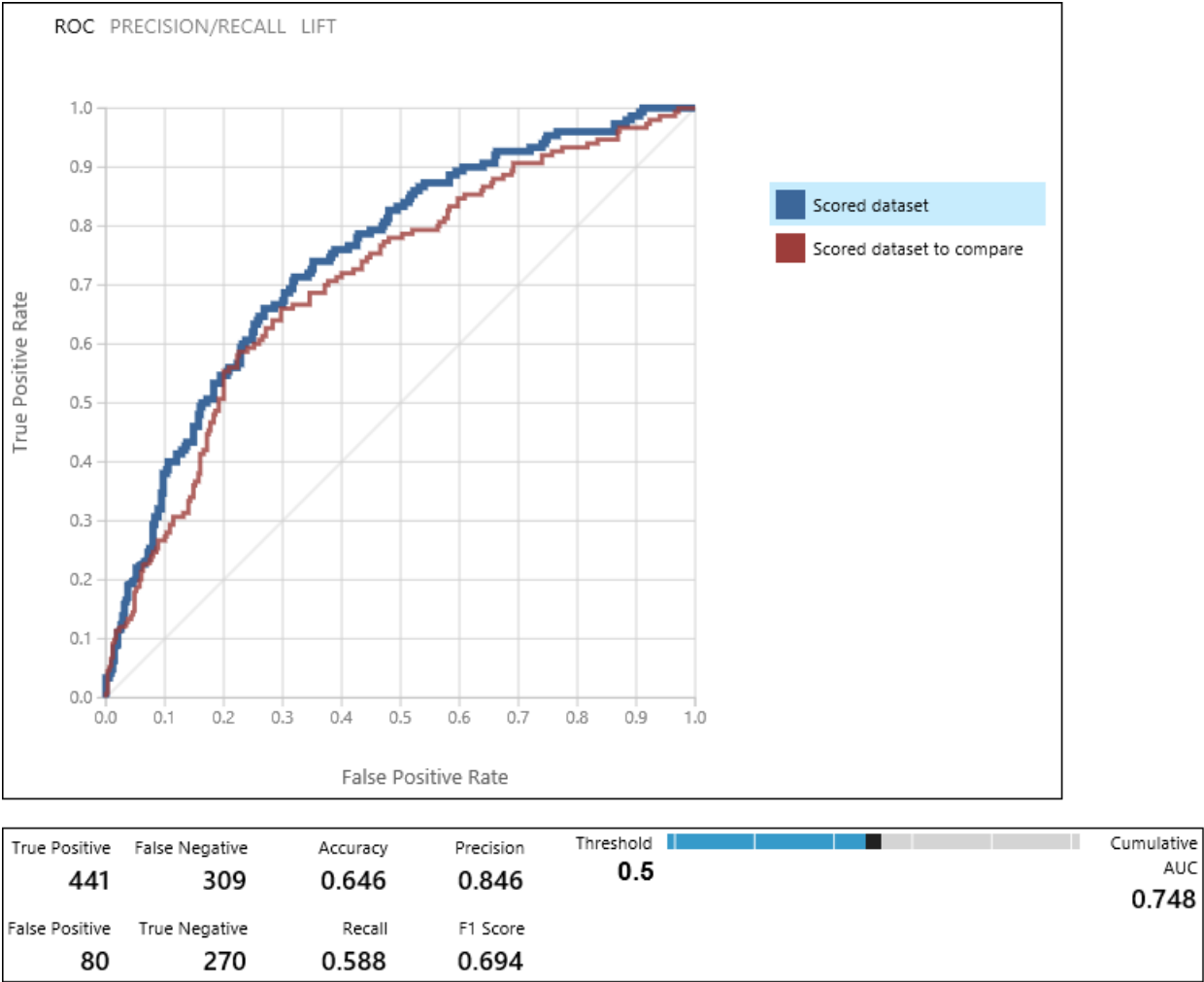


Чтобы проверить результаты, щелкните правой кнопкой мыши порт вывода модуля **Evaluate Model** (Оценка модели) и выберите **Visualize** (Визуализировать).

Модуль **Evaluate Model** (Оценка модели) создает пару кривых и метрики, которые позволяют сравнить результаты двух оцениваемых моделей. Результаты можно видеть в виде кривых ROC (рабочих характеристик приемника), кривых точности/полноты или кривых точности прогноза. Дополнительные отображаемые данные включают матрицу несоответствий, накопительные значения для области под кривой (AUC) и другие метрики. Вы можете изменить пороговое значение, перемещая ползунок влево или вправо, чтобы посмотреть, как это влияет на набор метрик.

В правой части графика щелкните **Scored dataset** (Оцененный набор данных) или

Scored dataset to compare (Оцененный набор данных для сравнения), чтобы выделить соответствующую кривую и отобразить внизу соответствующие метрики. В условных обозначениях для кривых "Scored dataset" (Оцененный набор данных) соответствует левому порту ввода модуля [Evaluate Model](#) (Оценка модели). В нашем случае это модель увеличивающегося дерева решений. "Оцененный набор данных для сравнения" соответствует правому входному порту — в нашем случае это модель SVM. Щелкните одну из этих меток, чтобы выделить кривую и отобразить внизу метрики для соответствующей модели, как показано на рисунке ниже.



Изучив эти значения, можно определить, какая модель обеспечивает более точные результаты. Можно вернуться назад и повторить эксперимент, изменяя значения параметров в моделях.

Наука и искусство интерпретации результатов и настройки производительности моделей выходят за рамки этого руководства. Для получения дополнительной информации ознакомьтесь со следующими статьями.

- [Как оценить эффективность модели в Студии машинного обучения \(классической\)](#)
- [Выбор параметров для оптимизации алгоритмов в Студии машинного обучения \(классическая модель\)](#)
- [Интерпретация результатов модели в Студии машинного обучения \(классической\)](#)

Совет

При каждом выполнении эксперимента запись об этой итерации сохраняется в журнале выполнения. Эти итерации можно просмотреть и вернуться к любой из них, нажав кнопку **ПРОСМОТР ЖУРНАЛА ВЫПОЛНЕНИЯ** под холстом. Можно также щелкнуть **Prior Run** (Предыдущее выполнение) на панели **Свойства**, чтобы открыть результаты предыдущей попытки сравнения.

Можно сделать копию любой итерации эксперимента, нажав кнопку **SAVE AS** (Сохранить как) под холстом. Чтобы отслеживать, какие данные вы сравнивали на каждом этапе выполнения эксперимента, используйте свойства эксперимента **Сводка** и **Описание**.

Дополнительные сведения см. в статье [Управление итерациями экспериментов в Студии машинного обучения \(классической\)](#).

Очистка ресурсов

Если вы не планируете дальше использовать ресурсы, созданные при работе с этой статьей, удалите их, чтобы плата не взималась. О том, как это сделать, см. в статье [Экспорт и удаление встроенных в продукт данных пользователей из Студии машинного обучения Azure](#).

Дальнейшие действия

В рамках этого руководства вы выполнили такие задачи:

- ✓ Создание эксперимента
- ✓ обучение нескольких моделей;
- ✓ Оценка и анализ моделей

Теперь вы готовы развернуть модели для этих данных.

Руководство. Часть 3. Классификация цветков ириса: развертывание модели