

RapidMiner

Одна Платформа. Делает все



Подготовка данных



Машинное обучение



Развертывание модели



студия

Визуальный рабочий процесс для
прогнозной аналитики

[УЧИТЬ БОЛЬШЕ](#)



Авто Модель

Создавайте прогнозирующие модели
быстрее

[УЧИТЬ БОЛЬШЕ](#)



Turbo Prep

Интуитивно понятная подготовка
данных

[УЧИТЬ БОЛЬШЕ](#)



сервер

Сотрудничество, управление и
развертывание



Подсчет очков в реальном
времени

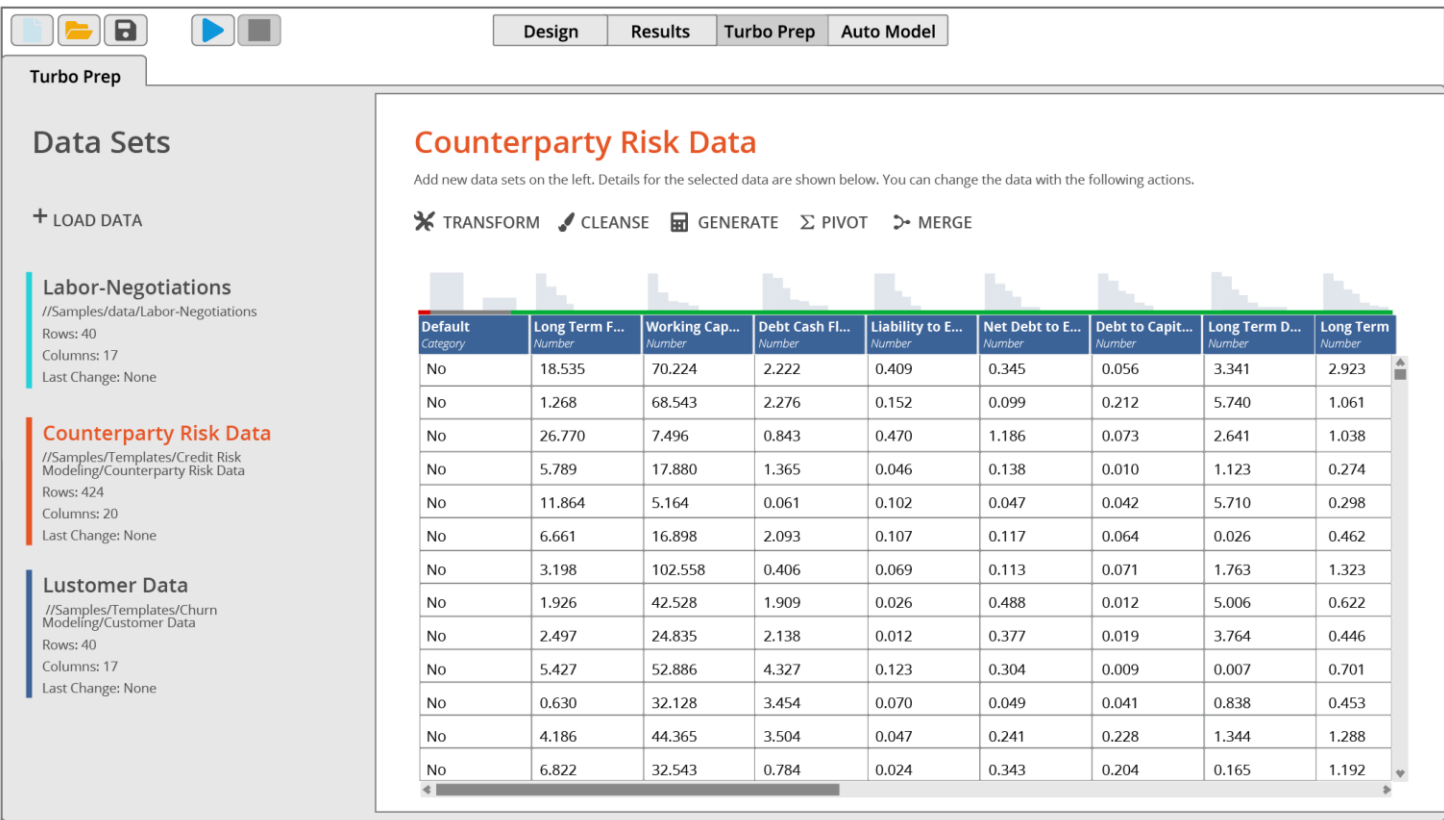
Превратите понимание в действие



Radoop

Код без данных для Hadoop и Spark

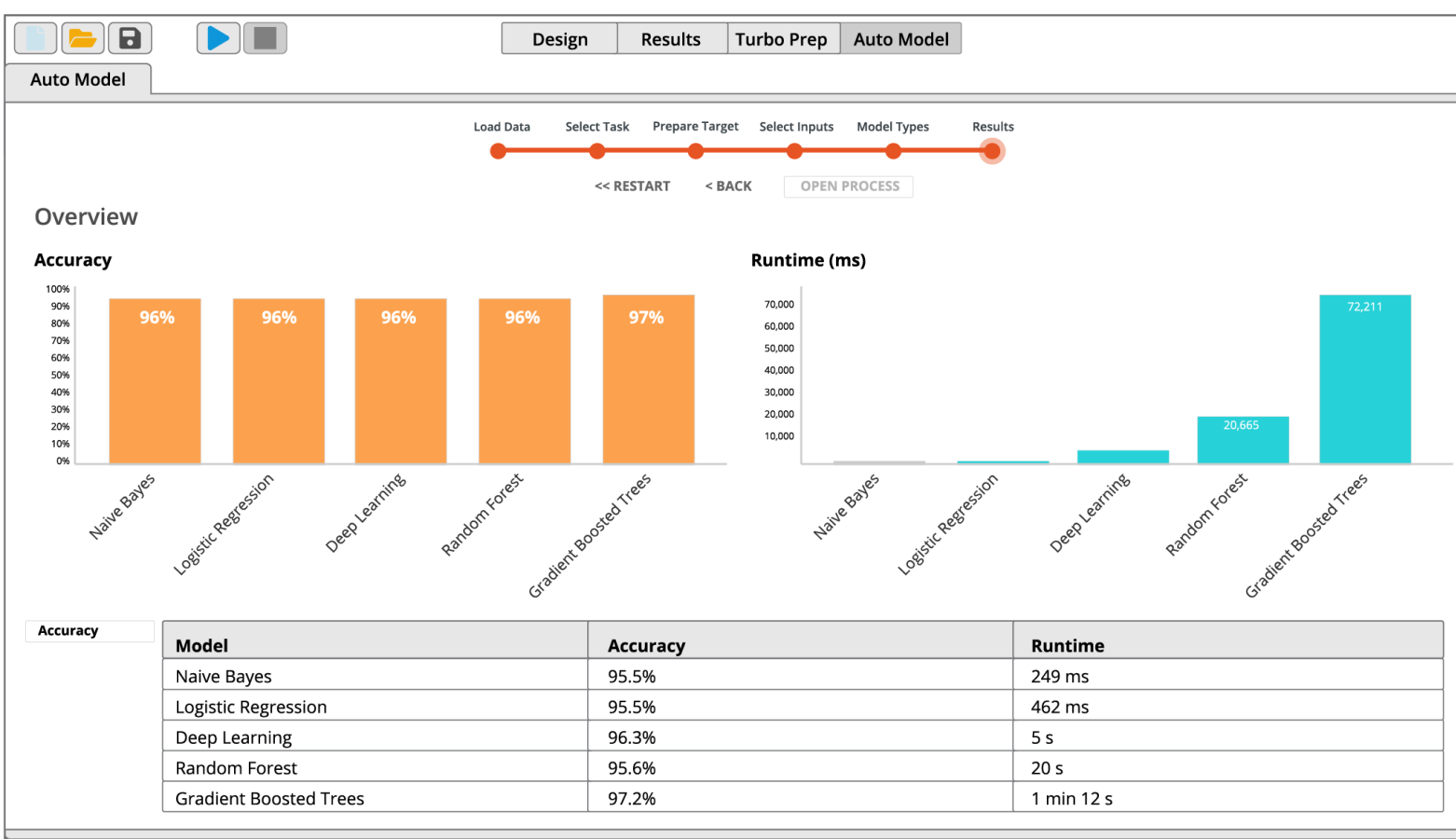
[УЧИТЬ БОЛЬШЕ](#)



Подготовка данных - это быстро, весело и интуитивно понятно

[RapidMiner Turbo Prep](#) позволяет легко подготовить данные для прогнозного моделирования. Интерактивно исследуйте данные, чтобы оценить их здоровье, полноту и качество. Быстро исправить общие проблемы, такие как пропущенные значения и выбросы. Смешайте несколько наборов данных вместе и создайте новые столбцы, используя простой редактор выражений.

Когда данные, наконец, будут готовы, создайте прогностические модели с помощью [RapidMiner Studio](#) и [Auto Model](#) или просто экспортируйте их в популярные бизнес-приложения, такие как Excel.



Создавайте прогнозные модели в 4 клика

[RapidMiner Auto Model](#) использует автоматическое машинное обучение и лучшие практики для построения прогностических моделей в 4 клика мышью. Просто укажите на данные, выберите модели для запуска и позвольте Auto Model использовать настройку гиперпараметров и автоматическую разработку функций для создания наилучшей модели.

Сохраните результаты Auto Model как процесс [RapidMiner Studio](#) для дальнейшей доработки и настройки, чтобы увидеть , как именно была создана модель.

Одна Платформа. Имеет ли *Everything* .

RapidMiner - это программная платформа для аналитических групп, **объединяющая подготовку данных, машинное обучение и развертывание прогностической модели** .



RapidMiner Studio

Визуальный дизайнер рабочих процессов для ученых данных



RapidMiner Server

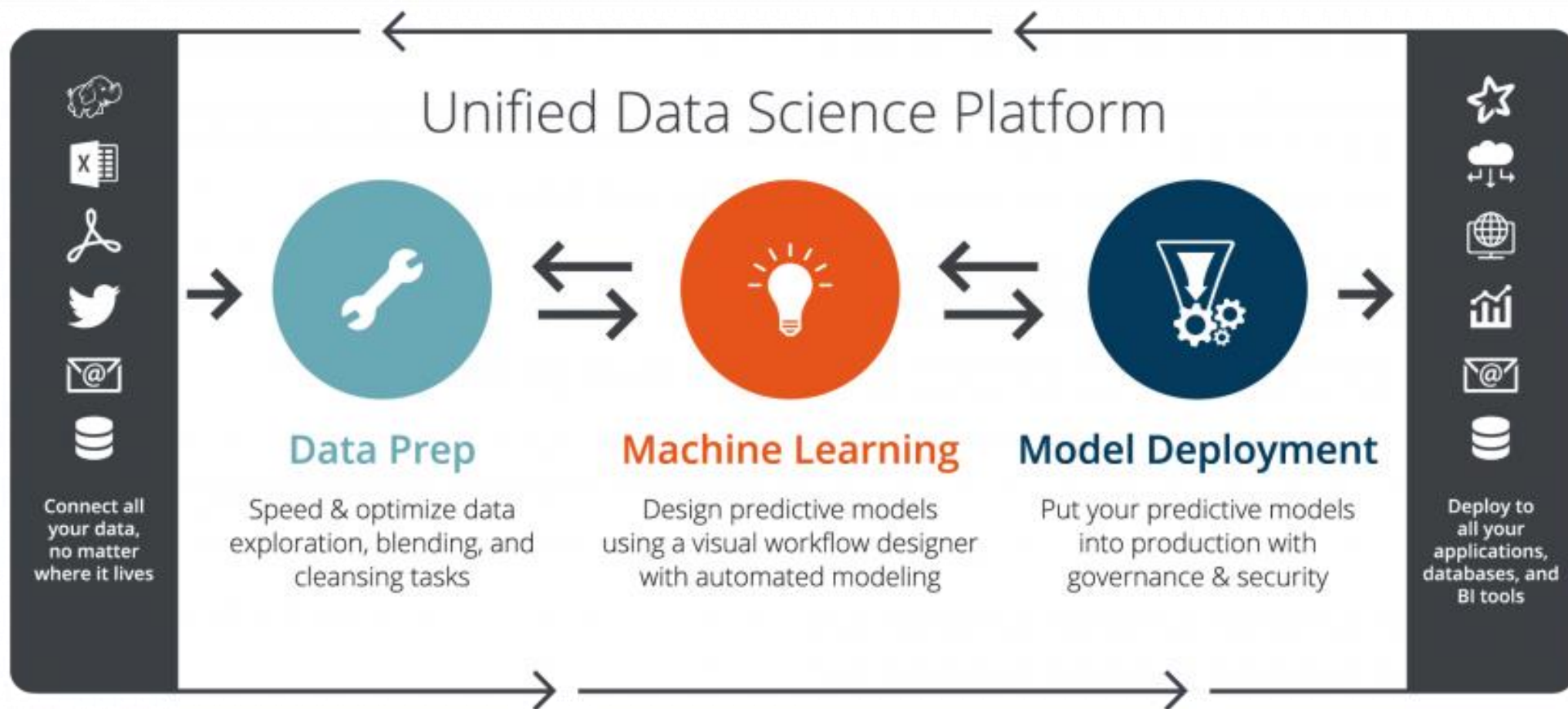
Совместная работа команды,
управление моделями и
развертывание



RapidMiner Radoop

Код без данных для Hadoop и Spark

Как это работает:



Комплексная платформа Data Science

- **Единая Платформа** . Одна платформа, один пользовательский интерфейс, одна система, поддерживают полный рабочий процесс от подготовки данных, развертывания модели до текущего управления моделью

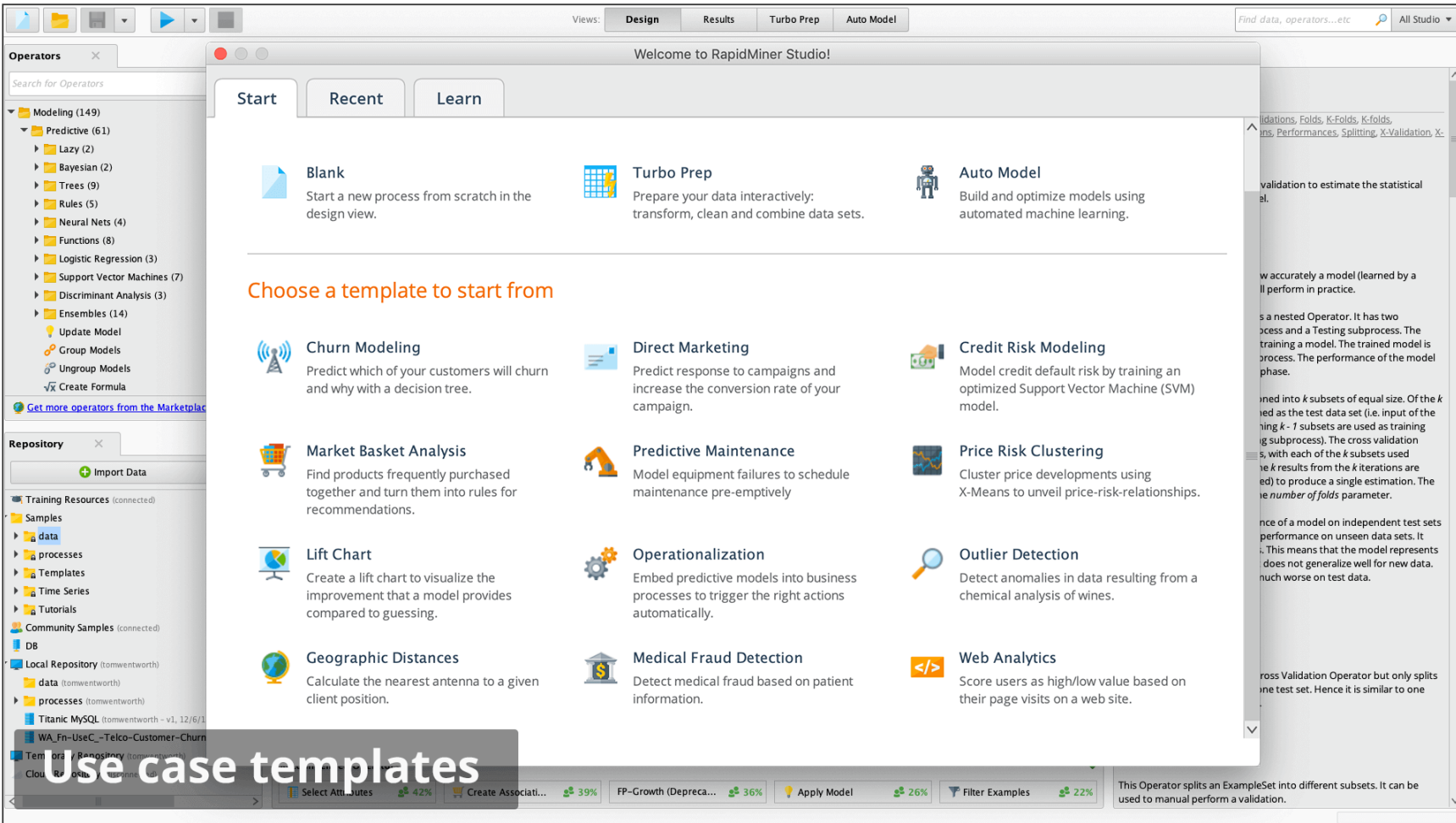
- **Визуальный дизайн рабочего процесса** . Быстрый в освоении и простой в использовании метод «перетаскивания» ускоряет сквозную науку о данных для повышения производительности

- **Широта функциональности** . Больше predefined функций машинного обучения и сторонних библиотек, чем на любой другой визуальной платформе

- **Инновации с открытым исходным кодом**. Хорошо принятые открытые языки и технологии, сообщество из более чем 250 000 экспертов по науке о данных и надежный рынок идут в ногу с растущими требованиями к науке о данных

- **Широкая связь** . Более 60 разъемов обеспечивают легкий доступ ко всем типам данных: структурированные, неструктурированные и большие данные

- **Наука данных в каждом масштабе**. Запускайте рабочие процессы в памяти или в Hadoop, предоставляя лучший вариант для проектов любого размера.

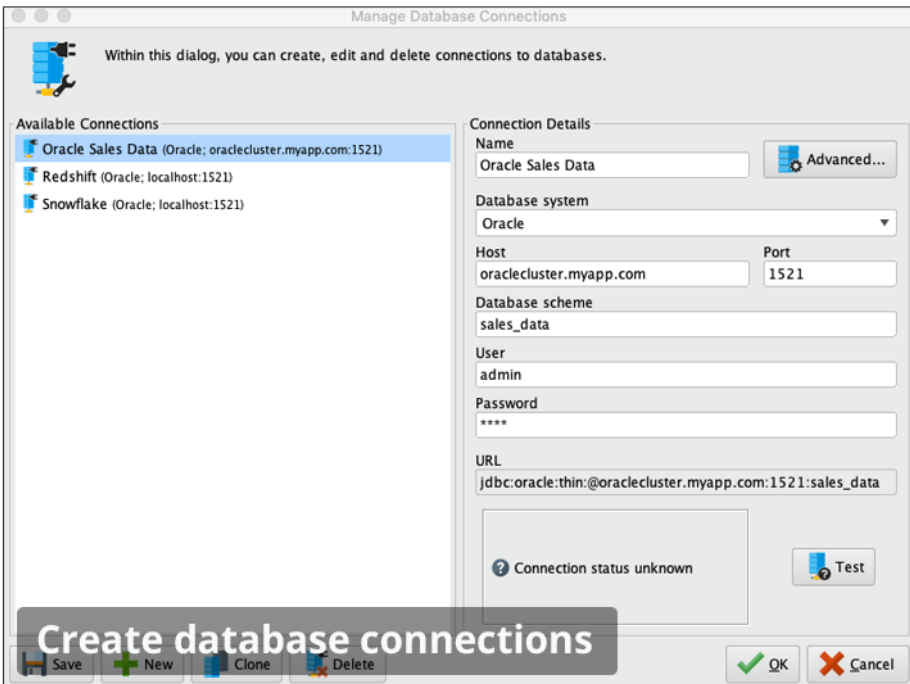


Visual Workflow Designer

Повысить
производительность всей
команды по науке о данных,
от аналитиков до экспертов

- Ускорьте создание прогнозирующих моделей, используя визуальный интерфейс drag + drop.

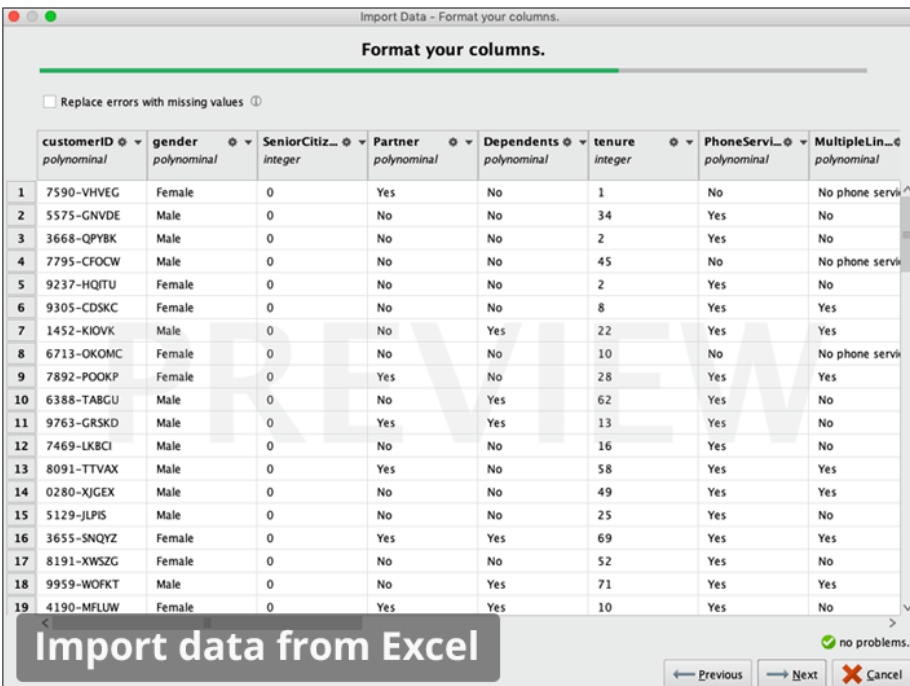
- Готовые шаблоны для общих случаев использования, включая отток клиентов, профилактическое обслуживание, обнаружение мошенничества и многое другое
- Уникальная функция Wisdom of Crowds обеспечивает проактивные рекомендации на каждом этапе рабочего процесса, включая набор параметров
- Богатая библиотека из более чем 1500 алгоритмов и функций машинного обучения для создания наилучшей модели для любого варианта использования

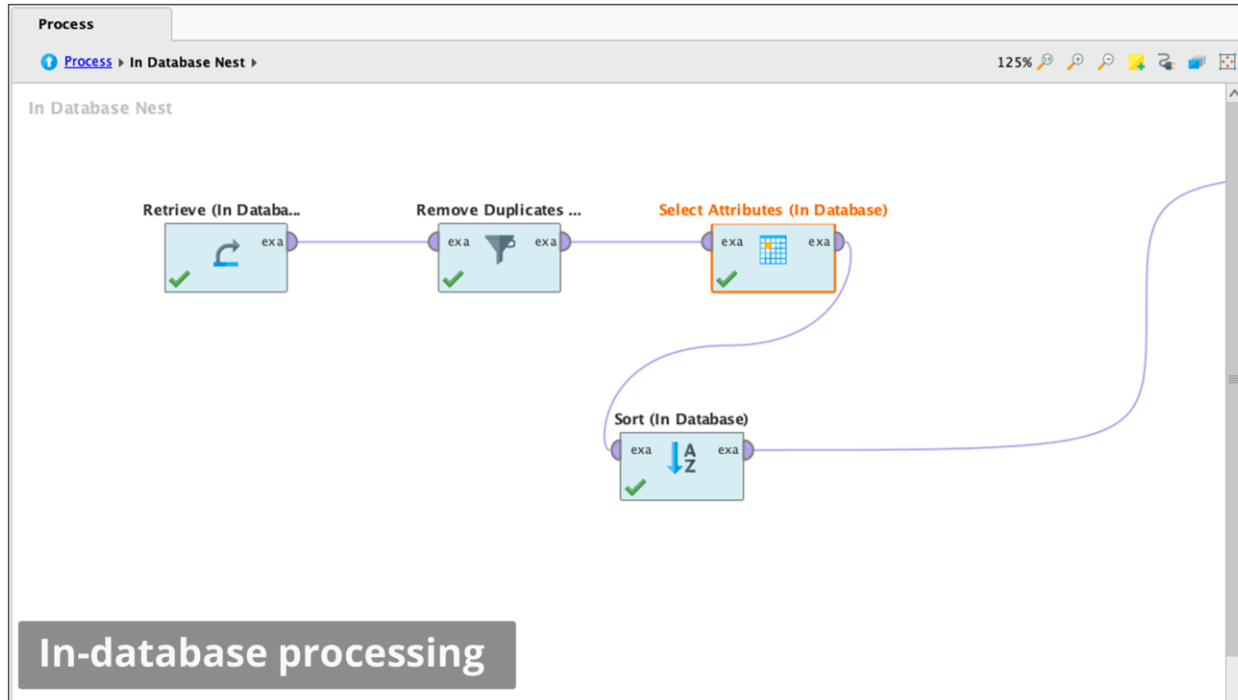


Источники данных

Подключайтесь ко всем своим данным, где бы они ни находились:

- Создавайте соединения по принципу point + click с базами данных, хранилищами, облачными источниками, документами, социальными сетями и бизнес-приложениями.
- Подключайтесь к новым источникам данных, загружая расширения с [RapidMiner Marketplace](#)

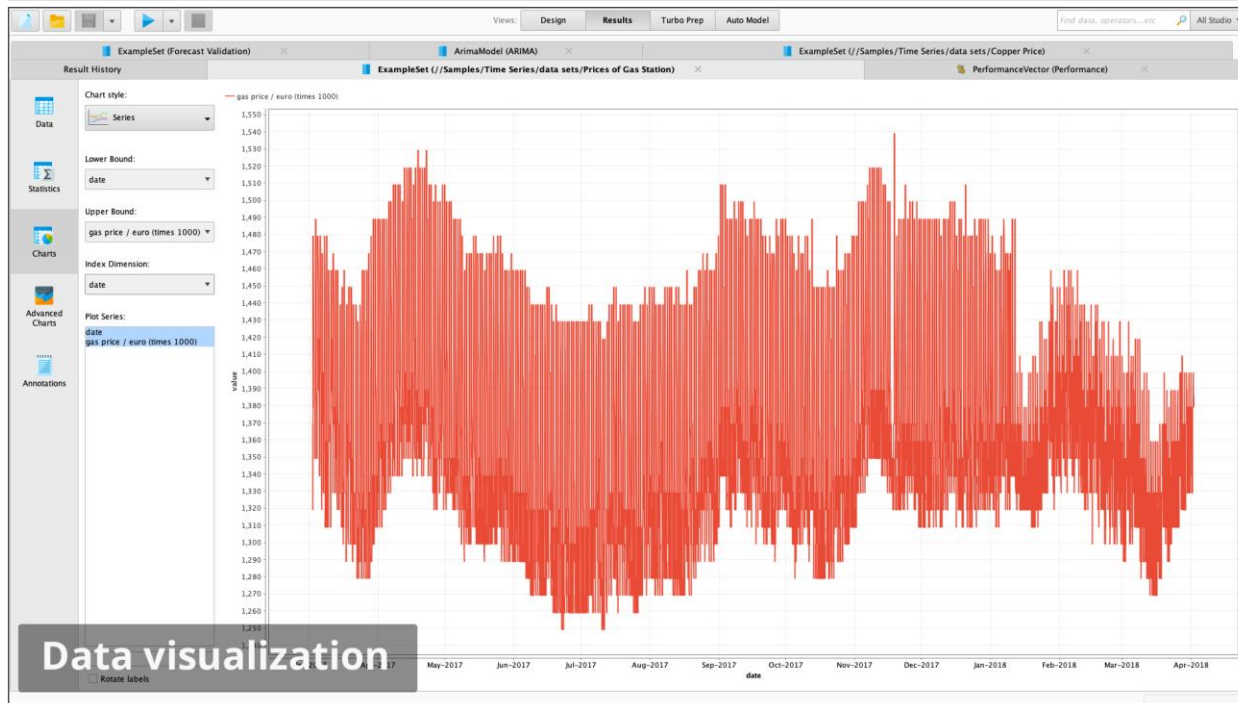




Обработка в базе данных

Запускать подготовку данных и процессы ETL непосредственно внутри баз данных

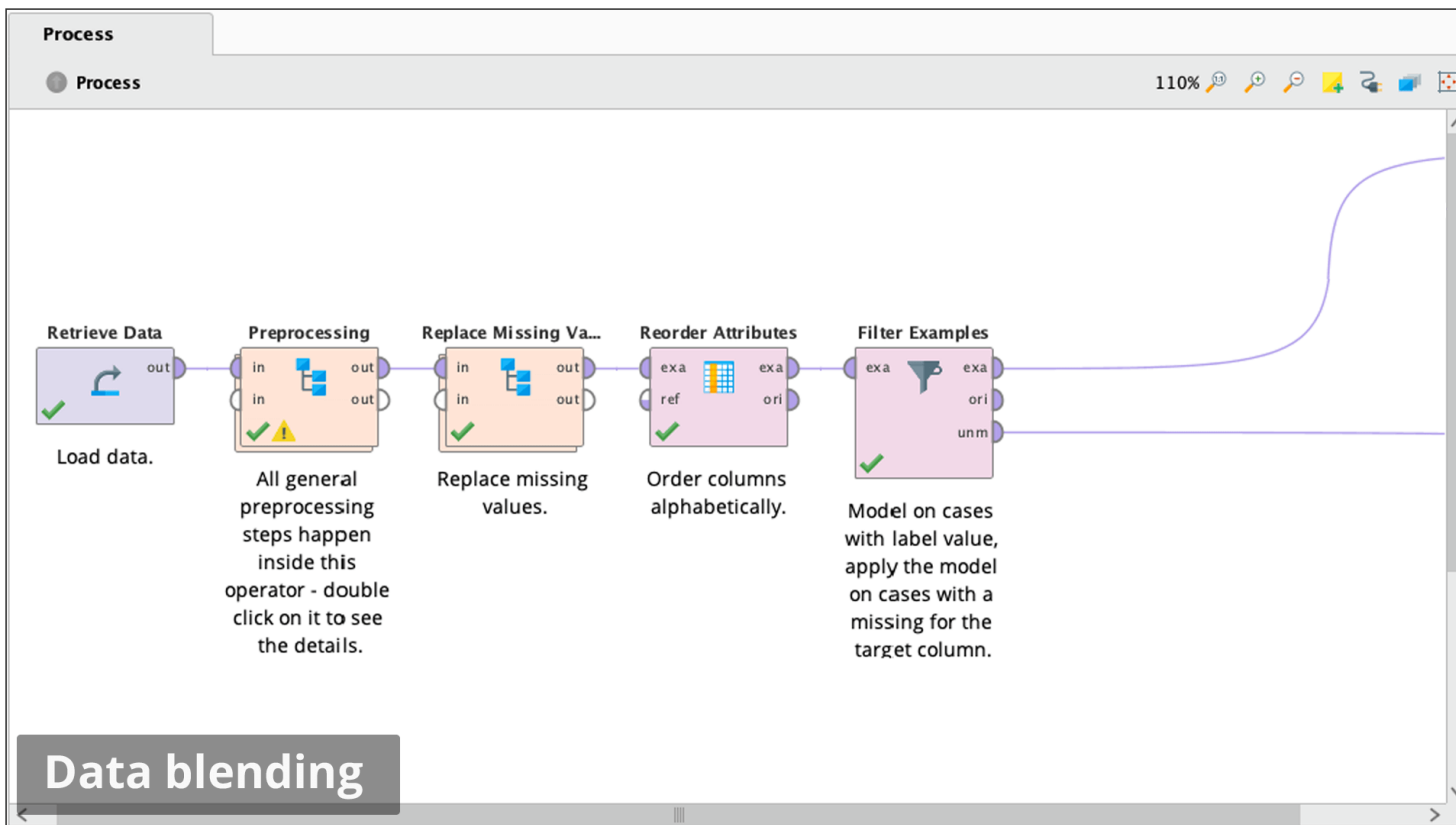
- Запрос и получение данных без написания сложного SQL
- Используйте мощные возможности масштабируемых кластеров баз данных
- Поддерживает MySQL, PostgreSQL и Google BigQuery



Исследуйте и визуализируйте данные

Оцените работоспособность, полноту и качество данных

- Исследуйте данные, используя надежные статистические обзоры и визуализации
- Выявление и устранение распространенных проблем с качеством данных, включая пропущенные значения и выбросы



Подготовка данных и смешивание

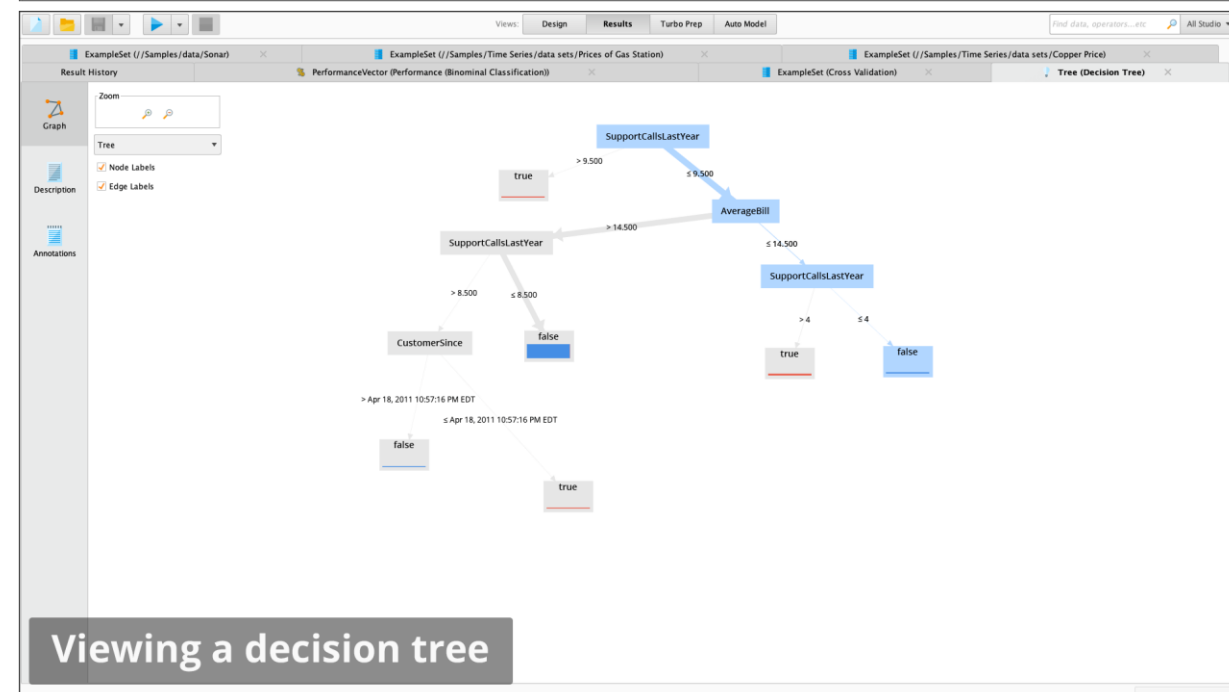
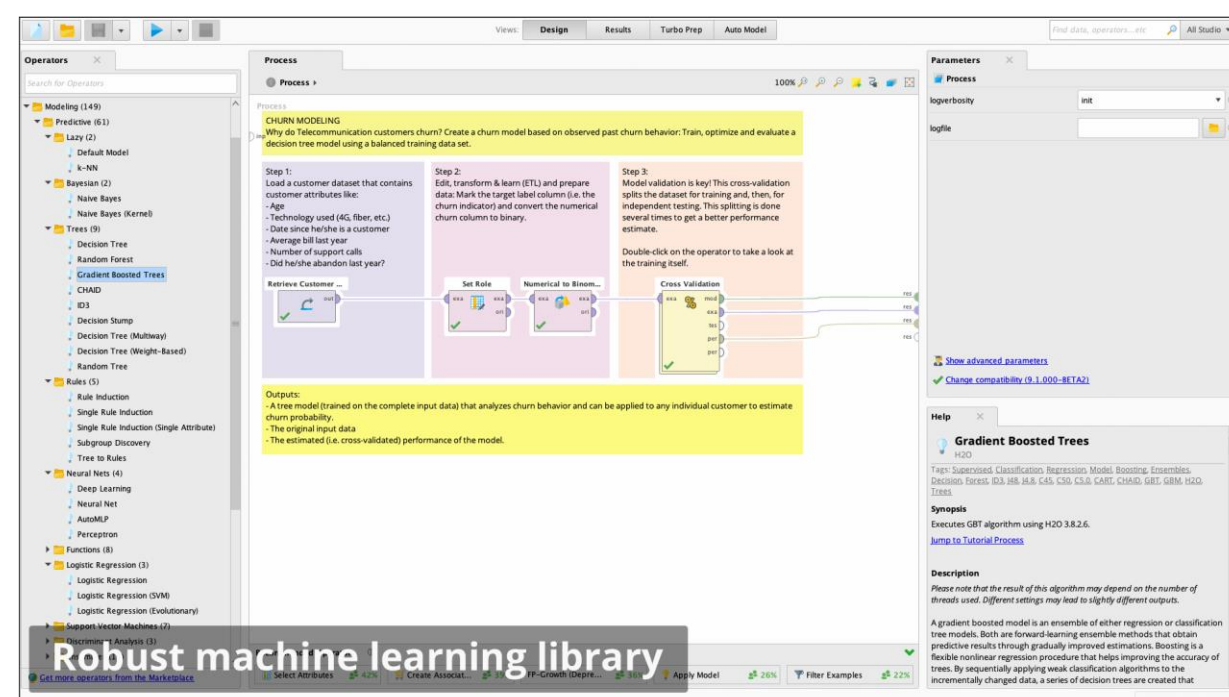
Устранить необходимость подготовки данных для прогнозного моделирования:

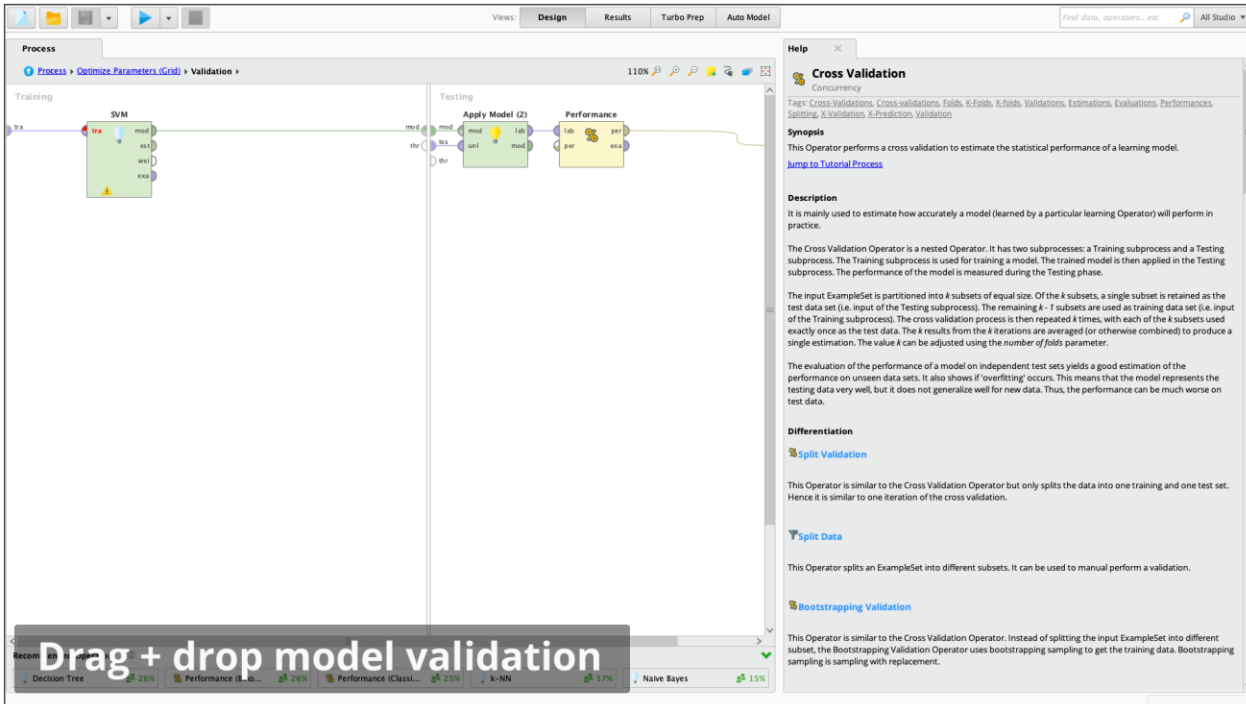
- Извлекать, объединять, фильтровать и группировать данные в любом количестве источников
- Создание повторяемых процессов подготовки данных и ETL, которые можно планировать и совместно использовать
- Перейдите к [RapidMiner Turbo Prep](#) для полностью интерактивной точки + опыт подготовки данных клика

Машинное обучение

Создавайте надежные модели машинного обучения без написания кода:

- Выберите из сотен контролируемых и неконтролируемых алгоритмов машинного обучения
- Реализация широкого спектра методов ML, включая регрессию, кластеризацию, временные ряды, анализ текста и глубокое обучение
- Используйте как автоматическое, так и ручное проектирование, чтобы повысить точность модели.
- Интеграция с [RapidMiner Auto Model](#) для создания моделей в 4 клика с использованием автоматического машинного обучения

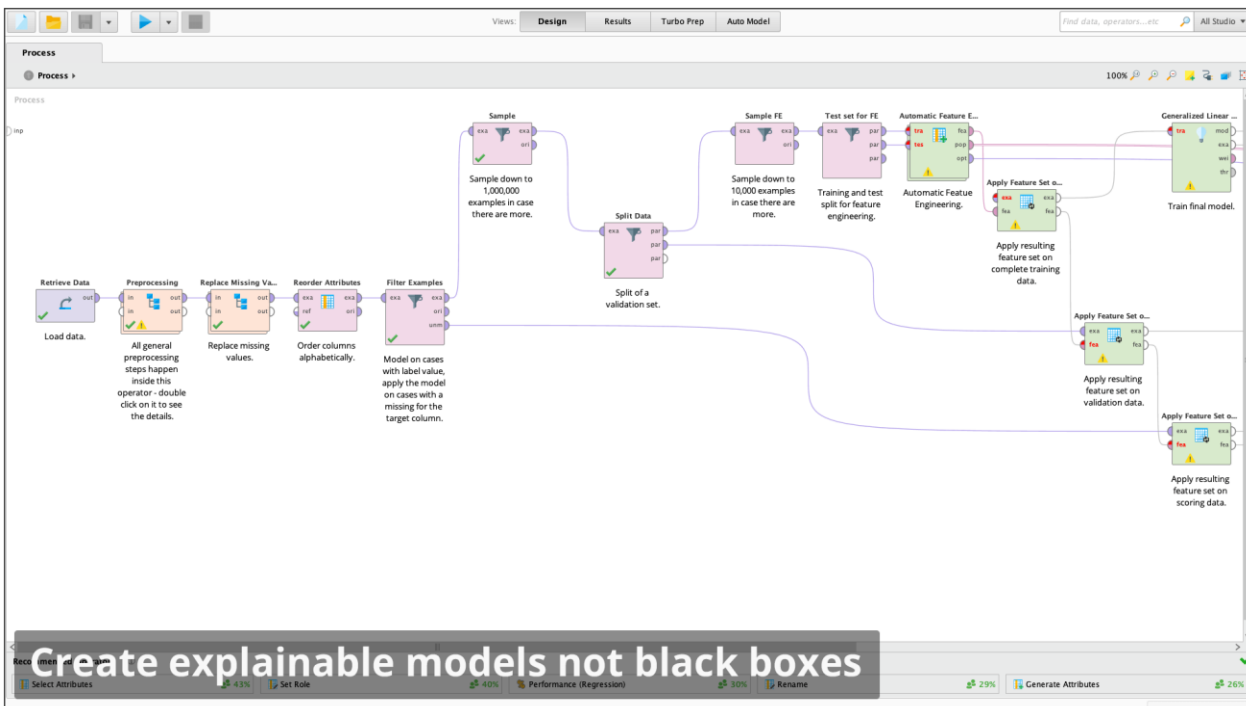




Проверка модели

Понять истинную производительность модели перед развертыванием в производство:

- Устраните переоснащение благодаря уникальному подходу, который предотвращает утечку данных предварительной подготовки модели в приложение модели
- Добавьте проверенные методы, такие как перекрестная проверка, в модель одним щелчком мыши



Объясняемые модели, а не черные ящики

Создавайте визуальные рабочие процессы в науке о данных, которые легко объяснить и легко понять:

- Каждый шаг в процессе подготовки, моделирования и проверки данных документирован для полной прозрачности
- Визуальный рабочий процесс легко объяснить другим в организации
- Поддерживает локальную интерпретируемую модель-Agnostic Explanations (LIME)

Гибкая оценка и развертывание

Превратите прогнозирующие идеи в предписывающие действия:

- Быстрое развертывание оцененных данных в электронных таблицах и средствах визуализации данных.
- Превратите модели в производственные веб-сервисы с RapidMiner Server
- Добавьте [RapidMiner Real-Time Scoring](#) для требовательных сценариев использования с высокой транзакцией / низкой задержкой

Открытый и расширяемый

Интеграция с существующими приложениями и кодом:

- Используйте существующий код R и Python и библиотеки для расширения RapidMiner
- Загрузите новые функциональные возможности через [RapidMiner Marketplace](#)
- Открытое ядро RapidMiner Studio доступно по лицензии GPL

The screenshot shows the RapidMiner interface with a decision tree model on the right and a spreadsheet on the left. The spreadsheet has columns for Row No., ChurnIndicator, Technology, Age, CustomerSince, SupportCalls, and AverageBill. A text box at the bottom of the spreadsheet reads "Deploy scored data to spreadsheets".

The screenshot shows the RapidMiner Marketplace website. The header includes the RapidMiner logo and a search bar. The main content area is titled "Get even more out of RapidMiner with Extensions." and lists various categories and top downloads. A footer section contains links for "Interested in sharing?", "Have questions?", and "Don't have RapidMiner?".

RapidMiner Авто Модель

Создание прогностической модели в 4 клика, используя передовые методы автоматизированного машинного обучения и обработки данных

Views: Design Results Turbo Prep Auto Model

Auto Model

Load Data Select Task Prepare Target Select Inputs Model Types Results

Selected: 18 / Total: 19

Deselect Red Deselect Yellow Select All Deselect All

Selec...	St...	Quality	Name	Correlati...	ID-ness	Stability	Missing
<input type="checkbox"/>	Red		Return on Equity	0.00%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Yellow		Long Term Financing of Working Capital	0.00%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Yellow		Expense to Revenue Ratio	47.50%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Working Capital Requirement	18.04%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Debt Cash Flow Coverage Ratio	12.12%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Liability to Equity	15.63%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Net Debt to Equity Ratio	9.90%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Debt to Capital Ratio	4.64%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Long Term Debt to Asset	10.01%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Long Term Debt to Tangible Asset	8.64%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Interest Coverage Ratio	6.08%	?	0.24%	0.00%
<input checked="" type="checkbox"/>	Green		Net Profit Margin	22.00%	?	0.24%	0.00%

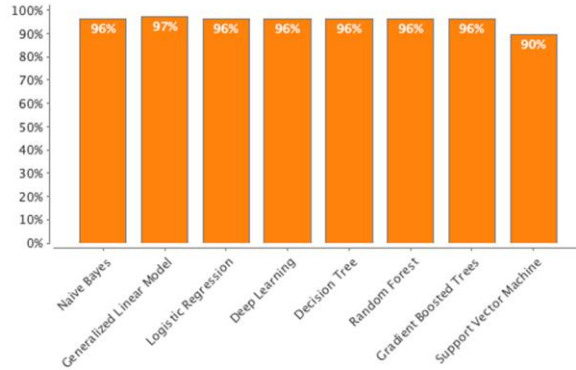
Guided data prep

Упрощенная подготовка данных для прогнозного моделирования

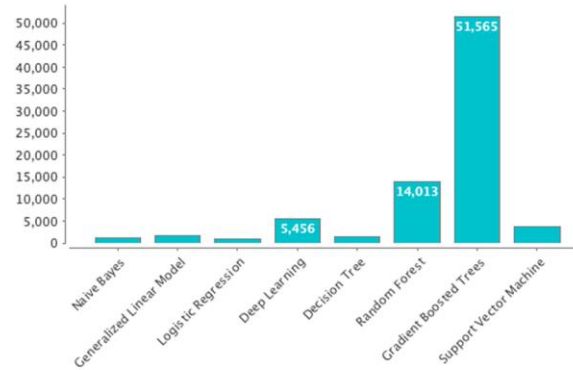
- Автоматический анализ данных для выявления общих проблем качества, таких как корреляции, пропущенные значения и стабильность
- Удалить проблемные столбцы данных одним щелчком мыши
- Дальнейшая подготовка, смешивание и обработка данных с помощью RapidMiner Turbo Prep

Overview

Accuracy



Runtime (ms)



Accuracy	Model	Accuracy	Runtime
	Naive Bayes	96.2%	1 s
	Generalized Linear Model	97.4%	2 s
	Logistic Regression	96.2%	1 s
	Deep Learning	96.2%	5 s

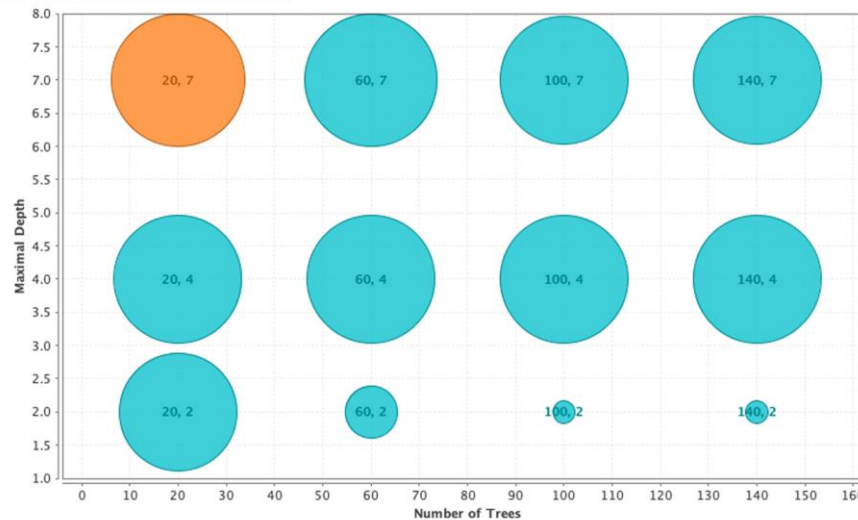
Find the best model

Random Forest – Optimal Parameters

Optimal Parameters

Number Of Trees: 20
Maximal Depth: 7

Performance for Parameters



Hyperparameter tuning

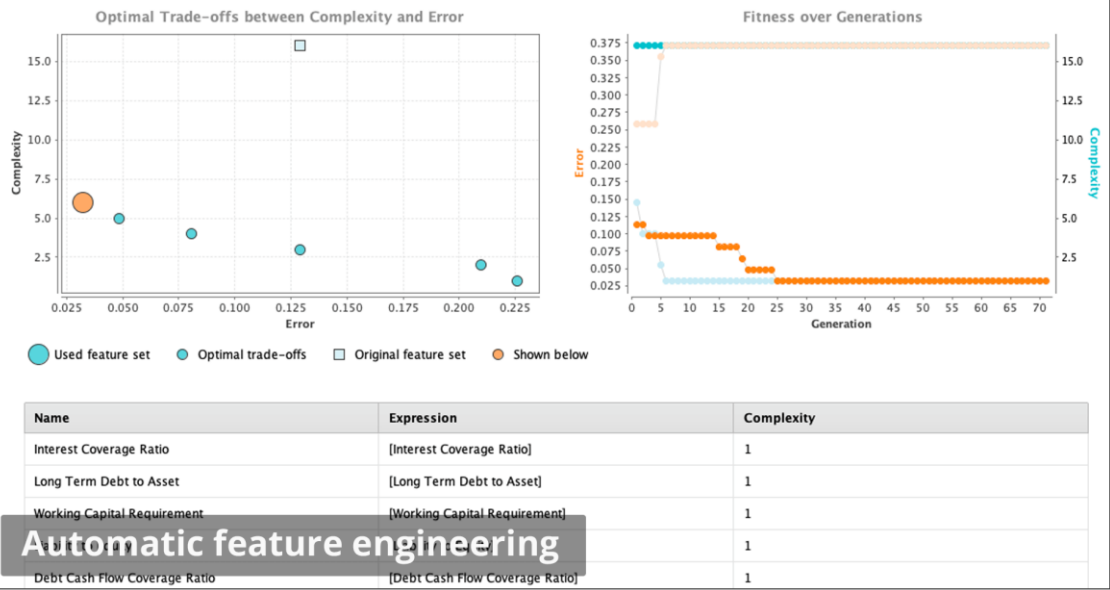
Number of Trees	Maximal Depth	Performance
20	2	0.926

Автоматический выбор модели и настройка

Найдите лучшую модель для ваших данных, используя несколько алгоритмов машинного обучения и оптимизацию гиперпараметров:

- Выберите из популярных моделей машинного обучения классификации, кластеризации и обнаружения выбросов
- Автоматическая настройка гиперпараметра находит лучшие параметры для модели
- Поддерживает собственные байесовские алгоритмы, обобщенную линейную модель, логистическую регрессию, глубокое обучение, дерево решений, случайный лес, деревья с градиентным ускорением (XGBoost), машины опорных векторов и кластеризацию k-Means

Generalized Linear Model – Feature Sets



Automatic feature engineering

Автоматическое проектирование

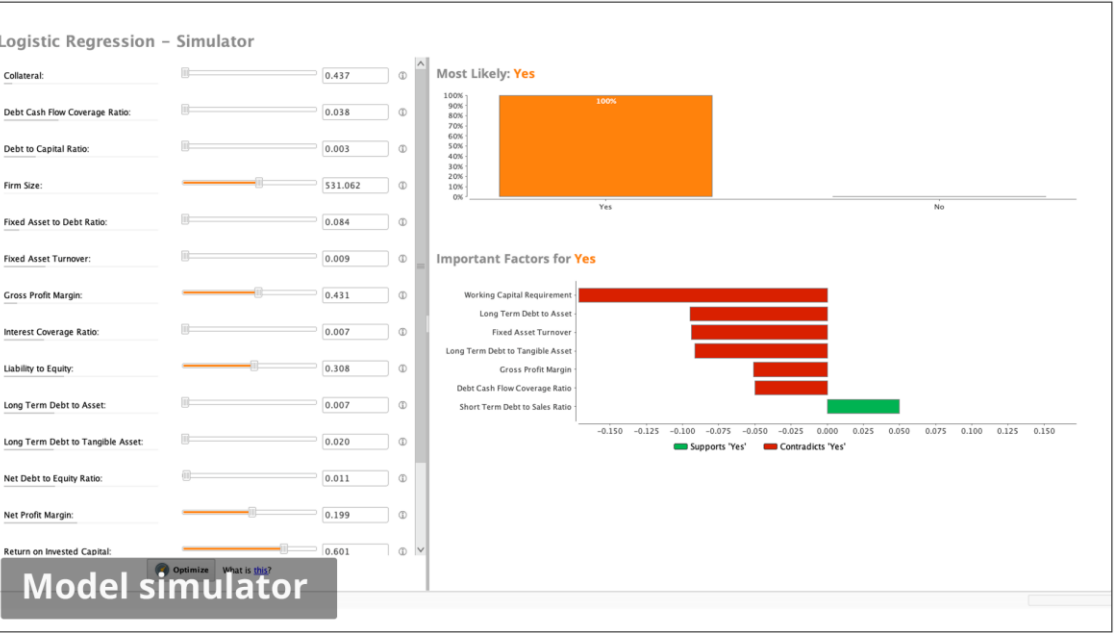
Автоматизированная разработка функций превращает хорошие модели в исключительные модели без ущерба для объяснимости

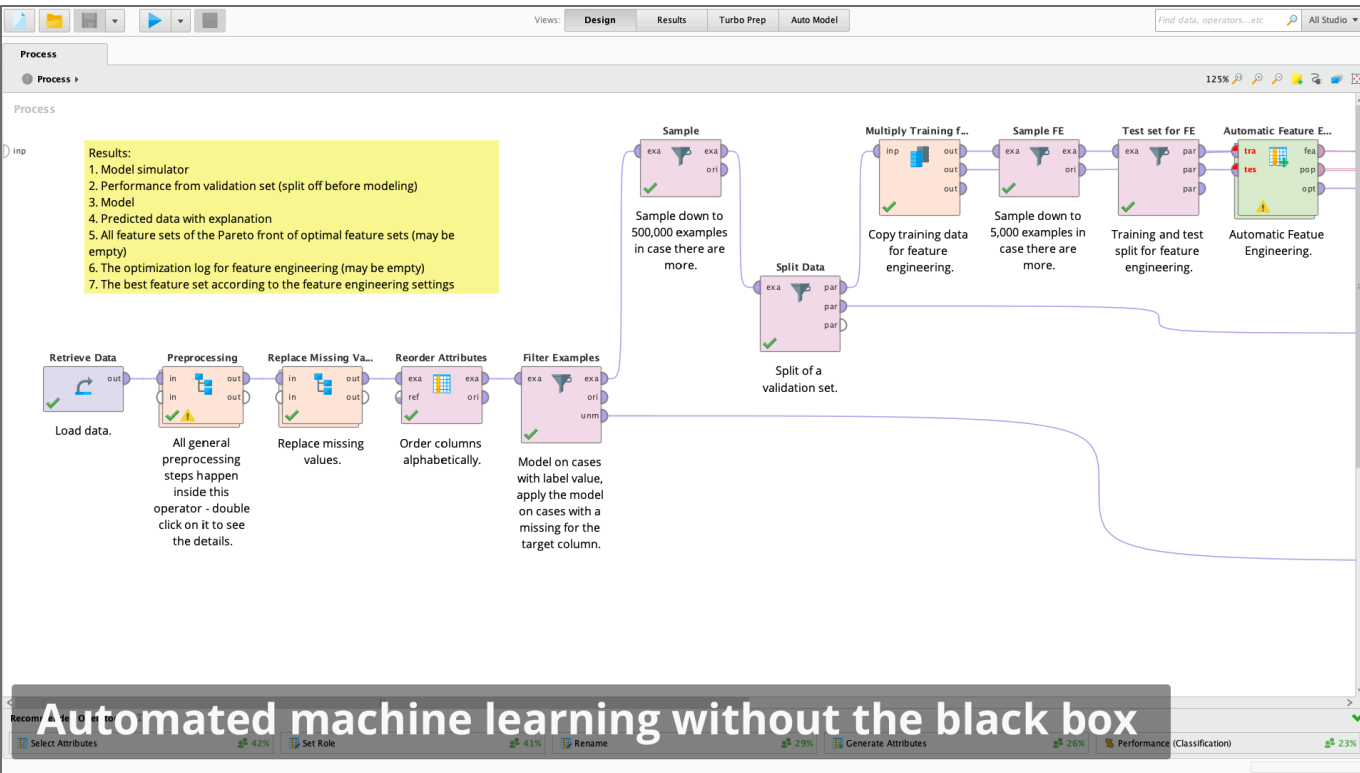
- Уникальный Evolutionary Feature Selection позволяет пользователям выбирать между точными, простыми и сбалансированными моделями
- Автоматически генерировать новые функции из данных для повышения точности модели
- Создавайте сотни тысяч моделей, чтобы найти лучшую

Превратите прогнозирующие модели в предписывающие действия

Принимайте лучшие бизнес-решения с помощью прогнозной аналитики

- Используйте простой интерфейс drag + drop, чтобы узнать, как изменения значений каждой функции влияют на производительность модели.
- Экспорт прогнозов в популярные бизнес-приложения, такие как Excel и инструменты визуализации данных





Нет черных ящиков

Создавайте прогностические модели, которые легко объяснить и которым можно доверять

- Посмотрите верхние переменные для каждой модели, чтобы увидеть, как именно модель пришла к своим выводам

- Прогнозирующая модель документируется и аннотируется шаг за шагом для полной прозрачности и доверия

- Каждый шаг в процессе рабочего процесса может быть настроен и улучшен в визуальном конструкторе рабочих процессов RapidMiner Studio

RapidMiner Turbo Prep

Смешивайте, обрабатывайте и очищайте данные с помощью интуитивно понятной подготовки данных, которую интересно использовать

Интуитивная подготовка данных

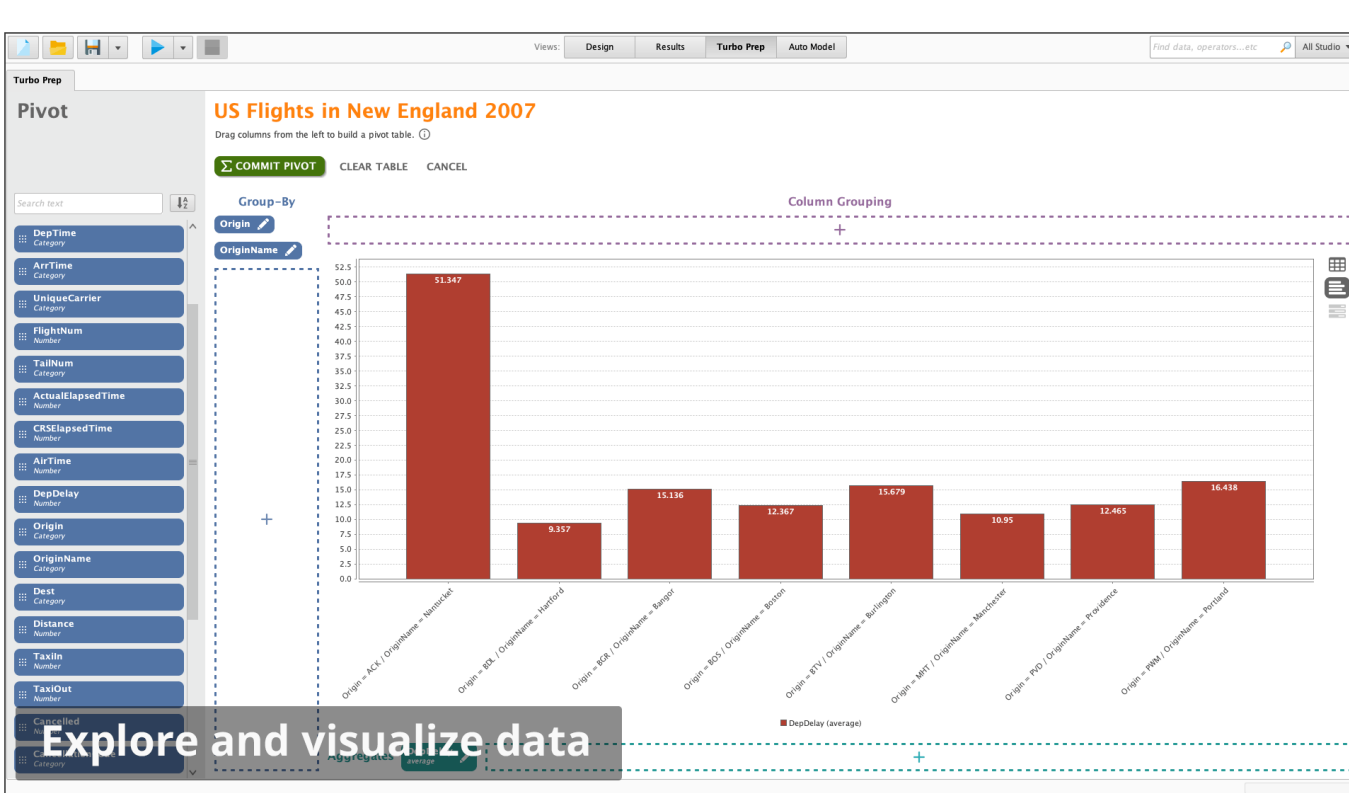
Работайте напрямую с данными в быстром, веселом и интерактивном режиме подготовки данных

- Внесите изменения и сразу увидите результаты, даже с массивными наборами данных
- Все изменения являются версионными и могут быть откатаны в один клик
- Передача данных между [RapidMiner Studio](#) и Turbo Prep

The screenshot displays the RapidMiner Turbo Prep interface. On the left, a sidebar lists available data sets: 'US Weather in New England 2007', 'US Flights in New England 2007', 'Sonar', and 'Titanic'. The main workspace shows the 'US Flights in New England 2007' dataset with a table of flight data. Above the table, there are tabs for 'Design', 'Results', 'Turbo Prep' (active), and 'Auto Model'. Below the tabs, there are buttons for 'LOAD DATA', 'TRANSFORM', 'CLEANSE', 'GENERATE', 'PIVOT', and 'MERGE'. The table has columns for Year, Month, DayofMonth, DayOfWeek, DepTime, ArrTime, UniqueCarrier, FlightNum, TailNum, ActualElapsedTime, CRSElapsedTime, AirTime, DepDelay, Origin, and OriginName. The bottom status bar indicates '227,674 rows - 29 columns (8 nominal, 19 numerical, 2 date)'.

Year	Month	DayofMonth	DayOfWeek	DepTime	ArrTime	UniqueCarrier	FlightNum	TailNum	ActualElapse...	CRSElapsedT...	AirTime	DepDelay	Origin	OriginN...
2007	1	2	2	712	832	WN	2049	N219	140	155	128	12	BDL	Hartfor
2007	1	2	2	749	850	WN	280	N772	61	70	51	-1	BDL	Hartfor
2007	1	2	2	1655	1755	WN	486	N208	60	70	52	25	BDL	Hartfor
2007	1	2	2	632	737	WN	711	N673	65	80	51	7	BDL	Hartfor
2007	1	2	2	1323	1430	WN	770	N344	67	70	55	13	BDL	Hartfor
2007	1	2	2	1814	1922	WN	805	N731	68	70	59	4	BDL	Hartfor
2007	1	2	2	1055	1155	WN	1047	N224	60	70	50	10	BDL	Hartfor
2007	1	2	2	2051	2152	WN	1292	N374SW	61	70	49	16	BDL	Hartfor
2007	1	2	2	1554	1656	WN	1603	N375	62	70	51	9	BDL	Hartfor
2007	1	2	2	827	927	WN	2043	N213	60	70	48	-3	BDL	Hartfor
2007	1	2	2	1458	1728	WN	267	N785	330	355	294	3	BDL	Hartfor
2007	1	2	2	1902	2147	WN	504	N748SW	165	175	154	17	BDL	Hartfor
2007	1	2	2	1007	1302	WN	1945	N475	175	185	160	-3	BDL	Hartfor
2007	1	2	2	1100	1401	WN	2285	N394SW	181	185	168	-5	BDL	Hartfor
2007	1	2	2	1758	1918	WN	145	N211	140	150	127	13	BDL	Hartfor
2007	1	2	2	650	813	WN	824	N412	143	150	131	10	BDL	Hartfor
2007	1	2	2	1244	1408	WN	1103	N435	144	150	130	-1	BDL	Hartfor
2007	1	2	2	718	826	WN	472	N496	68	75	56	3	BDL	Hartfor
2007	1	2	2	1603	1707	WN	2160	N386SW	64	90	49	-2	BDL	Hartfor
2007	1	2	2	1226	1338	WN	2320	N452	72	75	52	-4	BDL	Hartfor
2007	1	2	2	1925	2230	WN	1115	N678	185	195	170	0	BDL	Hartfor
2007	1	2	2	1330	1629	WN	2418	N728SW	179	190	167	0	BDL	Hartfor

Intuitive data prep



Explore and visualize data

Исследуйте и визуализируйте

Интерактивно исследуйте данные, чтобы оценить их здоровье, полноту и качество:

- Создание сводных таблиц для исследования и визуализации данных с использованием
- См надежные статистические обзоры и визуализации

Очистка

Радикально упростить очистку данных для машинного обучения:

- Используйте машинное обучение для автоматического удаления некачественных и коррелированных столбцов данных.
- Преобразование столбцов в числовые или категориальные в один клик
- Встроенные функции, такие как основной компонент и нормализация

Auto Cleansing

Define Target Improve Quality Change Types Handle Numbers Summary

RapidMiner can automatically perform common data cleansing techniques on your data to better prepare it for machine learning. In case you want to predict a column later on, please select it below.

No target column, thanks!

Passenger CL... Category	Sex Category	Age Number	No of Sibling... Number	No of Parents... Number	Passenger Fa... Number	Port of Emba... Category	Survived Category
First	Female	29	0	0	211.338	Southampton	Yes
First	Male	0.917	1	2	151.550	Southampton	Yes
First	Female	2	1	2	151.550	Southampton	No
First	Male	30	1	2	151.550	Southampton	No
First	Female	25	1	2	151.550	Southampton	No
First	Male	48	0	0	26.550	Southampton	Yes
First	Female	63	1	0	77.958	Southampton	Yes
First	Male	39	0	0	0	Southampton	No
First	Female	53	2	0	51.479	Southampton	Yes
First	Male	47	1	0	49.504	Cherbourg	No
First	Female	47	1	0	237.525	Cherbourg	No

Automated data cleansing

BACK NEXT

Merge

US Weather in New England 2007

Select a second data set to merge with on the left. Define how both data sets should be merged and see a preview below.

COMMIT MERGE CANCEL

UPDATE PREVIEW

ID Number	Max Gust Sp... Number	Precipitation... Number	Events Category	Weather Mo... Number	Weather Day Number	Max Temper... Number	Mean Temper... Number	Min Temper... Number	Dew PointC Number	MeanDew Po... Number	Min Dewpoint... Number	Min Humidity... Number	Max Visibilit... Number	Mean V... Number
1	66	29.720	Fog-Rain	1	1	12	6	-1	11	7	-3	73	16	6
2	58	5.840	Rain	1	2	10	7	3	9	3	-4	42	16	13
3	40	0	Normal	1	3	9	5	1	5	2	-4	60	16	16
4	38	0	Normal	1	4	11	9	7	8	7	4	79	16	14
5	38	0.760	Fog-Rain	1	5	12	10	8	11	9	7	88	16	7
6	52	4.570	Fog-Rain	1	6	13	11	10	12	11	7	77	16	3
7	42	0	Normal	1	7	10	7	4	6	0	-3	42	16	16
8	64	26.670	Fog-Rain	1	8	12	9	6	12	8	-1	62	16	7
9	52	0	Normal	1	9	7	4	3	-2	-4	-7	40	16	16
10	58	0	Normal	1	10	3	0	-2	-3	-9	-14	32	16	16
11	47	0	Normal	1	11	6	1	-4	0	-8	-13	44	16	16
12	47	0	Normal	1	12	10	8	7	5	4	2	68	16	15
13	42	4.570	Rain	1	13	10	8	6	9	7	4	86	16	9
14	32	0.250	Fog-Rain	1	14	8	7	6	8	6	3	85	16	8
15	38	0.510	Fog-Rain	1	15	10	8	6	9	7	5	93	16	5
16	45	1.520	Rain	1	16	11	3	-4	10	3	-12	53	16	10
17	55	0	Normal	1	17	-4	-7	-9	-11	-16	-19	35	16	16
18	47	0.760	Rain	1	18	7	-1	-9	6	-6	-14	61	16	15
19	58	10.670	Rain	1	19	8	4	2	7	1	-8	50	16	12
20	69	0	Normal	1	20	1	-2	-6	-8	-12	-17	37	16	16
21	60	0	Normal	1	21	-3	-5	-7	-14	-16	-17	36	16	16
22	16	0	Snow	1	22	1	-2	-4	-3	-8	-14	44	16	15

Columns from US Weather in New England 2007 Columns from US Flights in New England 2007 Join Keys 2,920 rows - 49 columns (10 nominal, 37 numerical, 2 date)

Data blending

Data blending

Объедините данные из нескольких источников, чтобы получить более унифицированное представление:

- Объедините несколько наборов данных всего несколькими щелчками мыши
- Машинное обучение автоматически определяет подходящие столбцы для объединения

Generate

US Weather in New England 2007

Generate a new column below. You can type a formula or drag in columns from the left and functions from the right. Update the preview and - if all is correct - commit the result.

COMMIT GENERATE CLEAR ALL CANCEL

Name: Flight Delayed

Formula:

```
if(Flight Delayed > 30, TRUE, FALSE)
```

UPDATE PREVIEW

Flight Delayed Category	DepDelay Number	ID Number	Max Gust Sp... Number	Precipitation... Number	Events Category	Weather Mo... Number	Weather Day Number	Max
false	12	1	66	29.720	Fog-Rain	1	1	12
false	-1	2	58	5.840	Rain	1	2	10
false	25	3	40	0	Normal	1	3	9
false	7	4	38	0	Normal	1	4	11
false	13	5	38	0.760	Fog-Rain	1	5	12
false	4	6	52	4.570	Fog-Rain	1	6	13
false	10	7	42	0	Normal	1	7	10
false	16	8	64	26.670	Fog-Rain	1	8	12
false	9	9	52	0	Normal	1	9	7
false	10	10	58	0	Normal	1	10	3

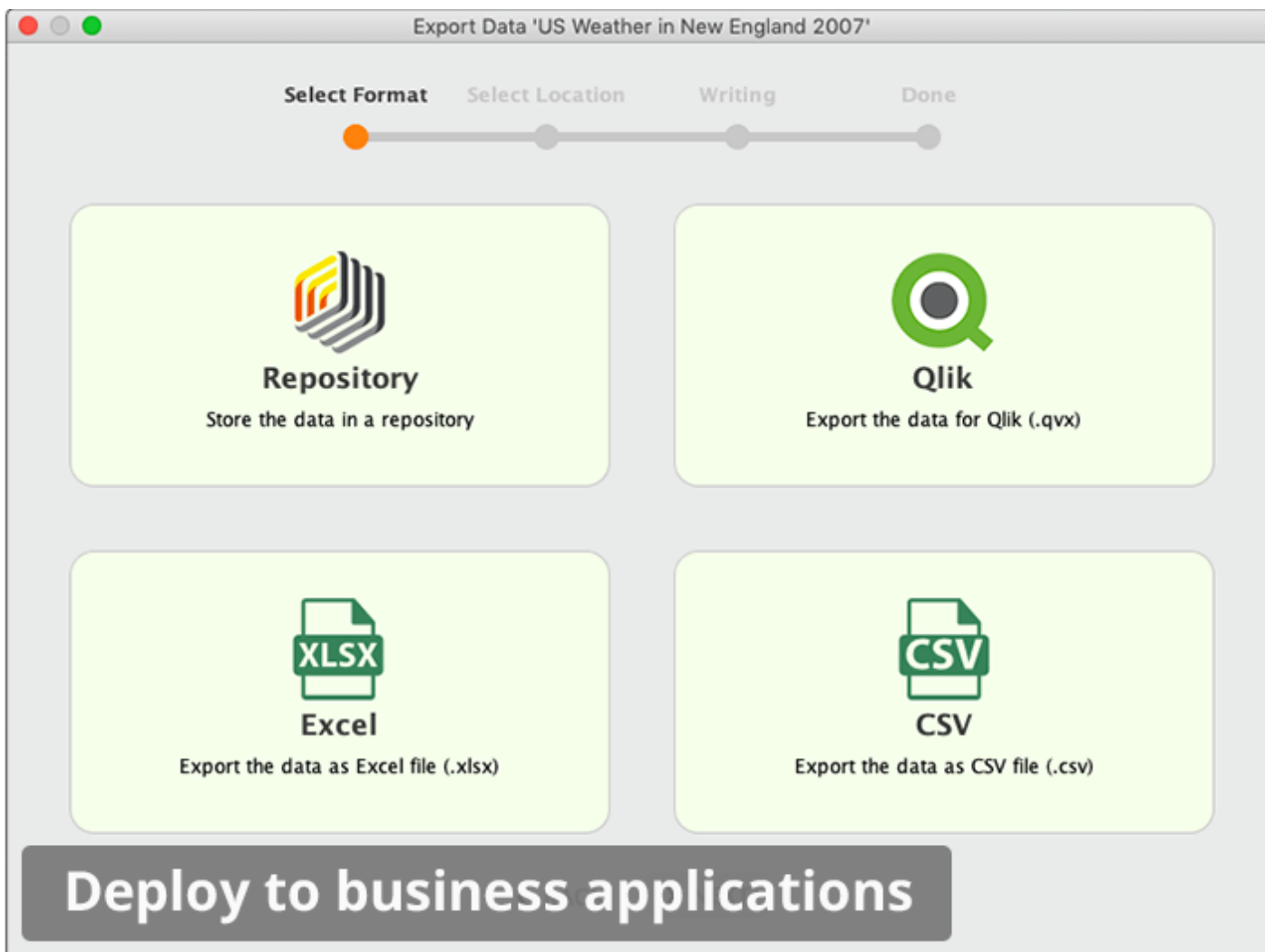
Used Column Generated Column 2,920 rows - 59 columns (11 nominal, 37 numerical, 2 date)

Data wrangling

Wrangle

Подготовьте данные для прогнозного моделирования и анализа:

- Извлекать, фильтровать, преобразовывать и группировать данные
- Создавайте новые столбцы данных, используя простой, но мощный редактор выражений



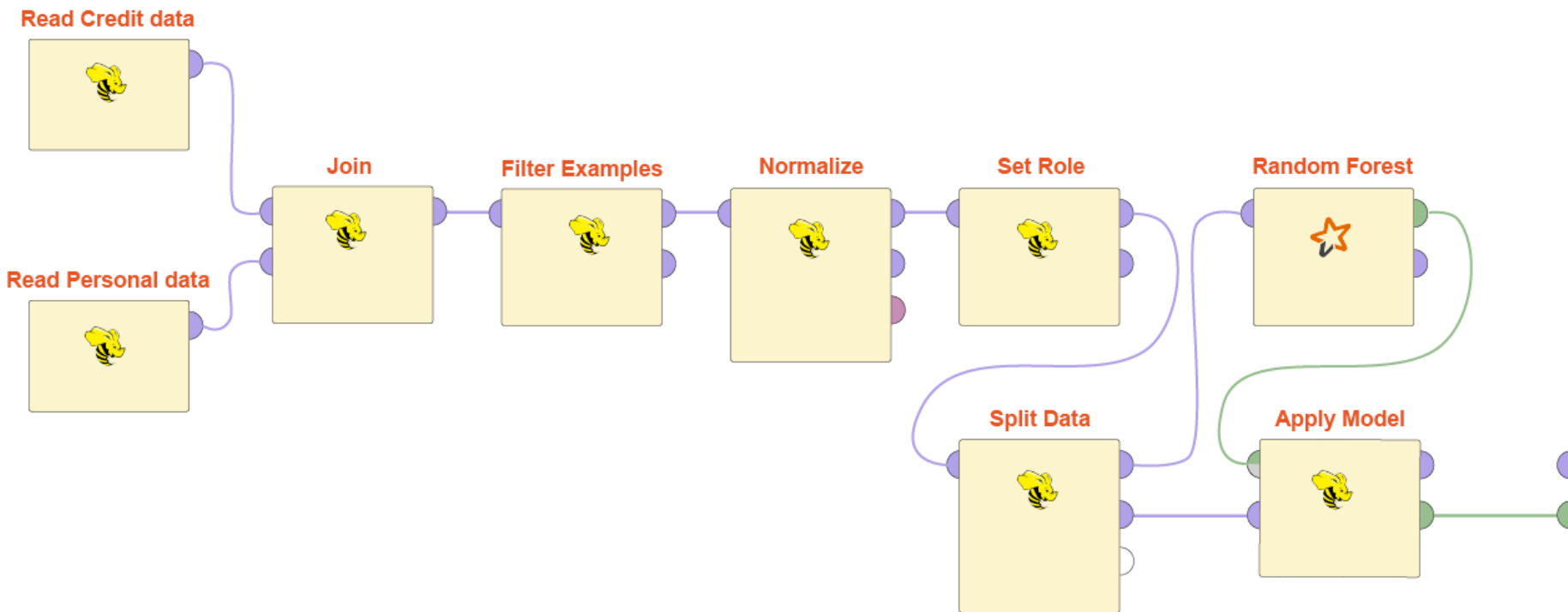
Повторно использовать и поделиться

Подключите данные к бизнес-процессам, чтобы превратить понимание в действия:

- Отправка данных в [RapidMiner Studio](#) или [RapidMiner Auto Model](#) для прогнозного моделирования
- Экспорт данных в электронные таблицы и инструменты визуализации данных
- Создание повторяемых процессов подготовки данных и ETL, которые можно планировать и совместно использовать

RapidMiner Radoop

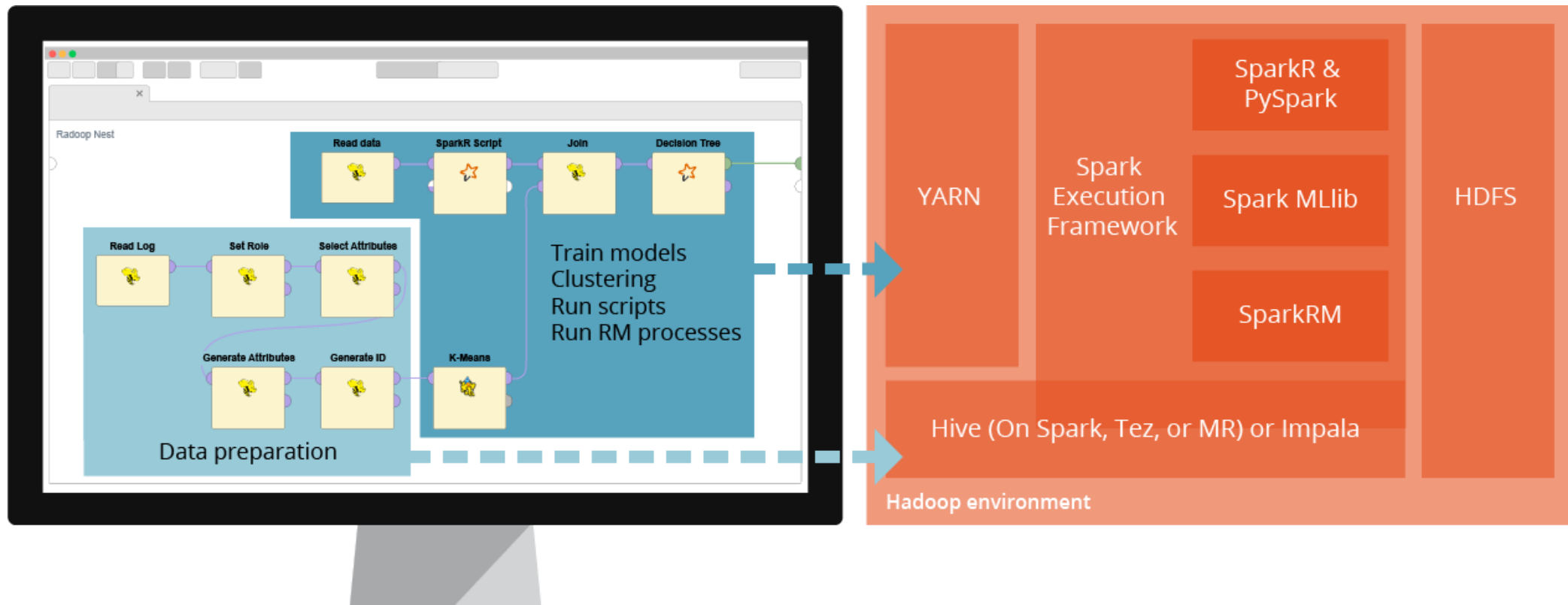
Устранить сложность Data Science на Hadoop и Spark



Машинное обучение без кода для Hadoop & Spark

Создавайте и запускайте прогностические модели в Hadoop, не кодируя в Spark.

- Создавайте прогнозные модели с помощью визуального дизайнера рабочих процессов RapidMiner Studio
- Расширение за пределы MLlib для решений более широкого набора вариантов использования, включая временные ряды и анализ текста
- Поддерживает HDP 3, Cloudera 6 и MapR 6



Используйте всю мощь кластеров Hadoop

Выполнять задания по подготовке данных и машинному обучению непосредственно в Hadoop

RapidMiner SparkRM позволяет параллельно выполнять все операции и потоки данных в RapidMiner Studio внутри Hadoop

Рабочие места автоматически переводятся на Spark и Hive

Никакого дополнительного программного обеспечения не требуется в кластерной среде Hadoop



Поддерживает стандарты и безопасность Hadoop

Максимизируйте свои инвестиции в экосистему Hadoop:

- Повторно использовать существующий код SparkR, PySpark, Pig и HiveQL
- Снижение риска и обеспечение соответствия нормативным требованиям благодаря встроенной поддержке Apache Sentry & Apache Ranger
- Развертывание шифрования HDFS для соблюдения политик безопасности данных