



Глазков Артем, консультант SAS Россия / СНГ

Управление моделями с помощью SAS Model Manager

27 июня 2019 | Алматы

Практика использования машинного обучения

ожидание - реальность

Что предлагает
Open Source...



Модель
(алгоритм)

Что требуется
дополнительно...



Встраивание в
бизнес

Занимайтесь машинным обучением а не кодом-обвязкой



[Google, Inc.: Hidden Technical Debt in Machine Learning Systems](#)

Что получает бизнес при использовании SAS Viya модуль SAS Model Manager



создание моделей



встраивание в бизнес

Как эффективно встроить любую модель в бизнес с помощью SAS Model Manager



Создание модели

Регистрация модели

Сравнение моделей

Выбор чемпиона

Валидация чемпиона

Публикация модели

Как эффективно встроить любую модель в бизнес с помощью SAS Model Manager



Создание модели

Регистрация модели

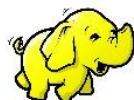
Сравнение моделей

Выбор чемпиона

Скоринг

Валидация чемпиона

{ REST }



Публикация модели



Скоринг новых данных

модели можно применять по разному

В режиме Real-time:

In-Stream (ESP)

In-MAS

In-Spark

В пакетном режиме:

In-SAS

In-DB

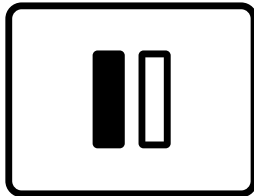
In-Hadoop

Как эффективно встроить любую модель в бизнес с помощью SAS Model Manager



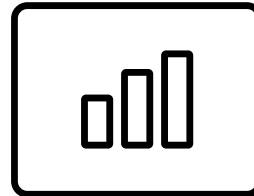
Мониторинг: встроенная библиотека отчетности с возможностью создавать собственные

классификация



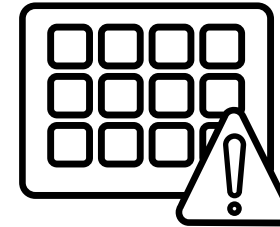
Variable distribution,
Deviation index,
Lift, ROC, Gini,
KS, Event/Non-Event Rate

регрессия



Variable distribution,
Deviation index,
ASE

пользовательские



Quantiles difference,
ROC AUC difference,
T-test,
seasonal T-test

+ создаваемые пользователем

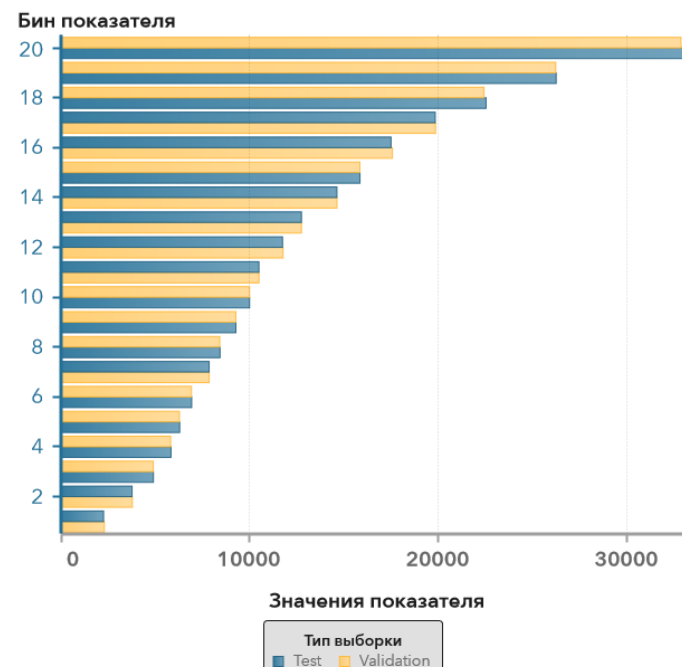
Пример пользовательского отчета интерактивная проверка статистических гипотез по T-тест

Filters: loan_amnt > ГИПОТЕЗА: Распределения тестовой и валидационной выборк...

Для старта выберите интервальную переменную:

loan_amnt

Распределение фактора в тестовой и валидационной выборке



Значения статистик в выборках данных

Тип выборки	Первая квартиль	Третья квартиль
Test	7,000.0	17,000.0
Validation	6,400.0	16,000.0

Значения отклонений по статистикам

Отклонение по 1-й квартили	Отклонения по 3-й квартили
9.4%	6.3%

Укажите значение порога значимости критерия

0.03
0.0% 100.0%

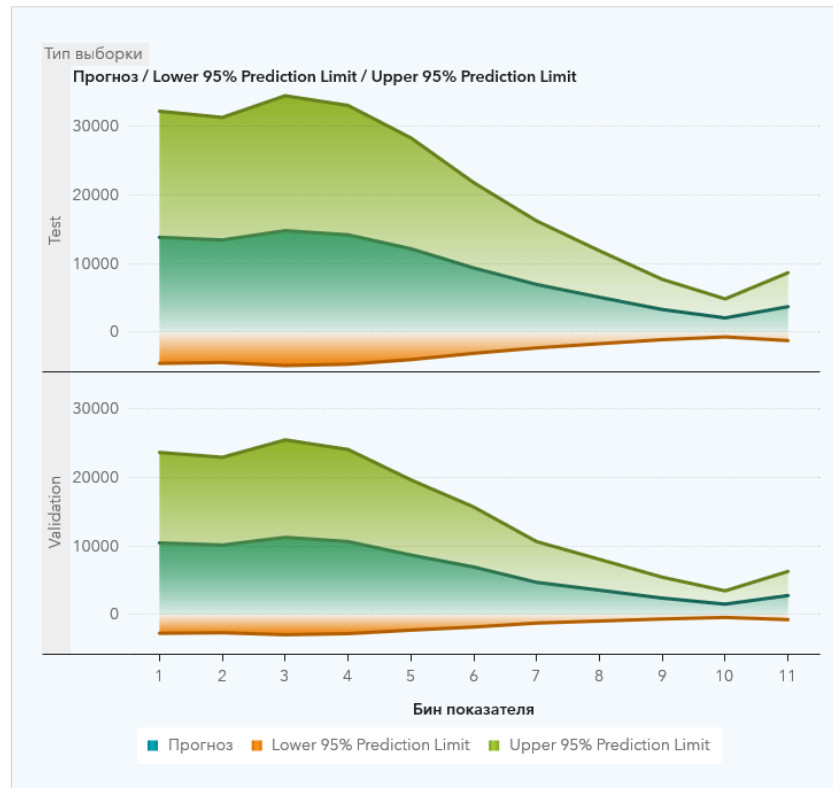
РЕЗУЛЬТАТ ПРОВЕРКИ ГИПОТЕЗЫ:

Различия статистически значимы

Пример пользовательского отчета проверка по Т-тест с учетом сезонности

Для старта выберите интервальную переменную:

acc_open_past_24mths



Оценивается смещение среднего с учетом сезонности

< ГИПОТЕЗА: T-test остатков относительно тестовой и валидации >

Equality

Variances	Method	DF	Pr > t	t Value
Equal	Pooled	40203	12E-21	9.32
Unequal	Satterthwaite	35742	11E-21	9.33

Ttest

Method ▲	Denominator Degr...	F Value	Numerator Degrees...	Pr > F
Folded F	16572	1.01	23631	0.6291

< ГИПОТЕЗА: T-test относительно тестовой и валидации >

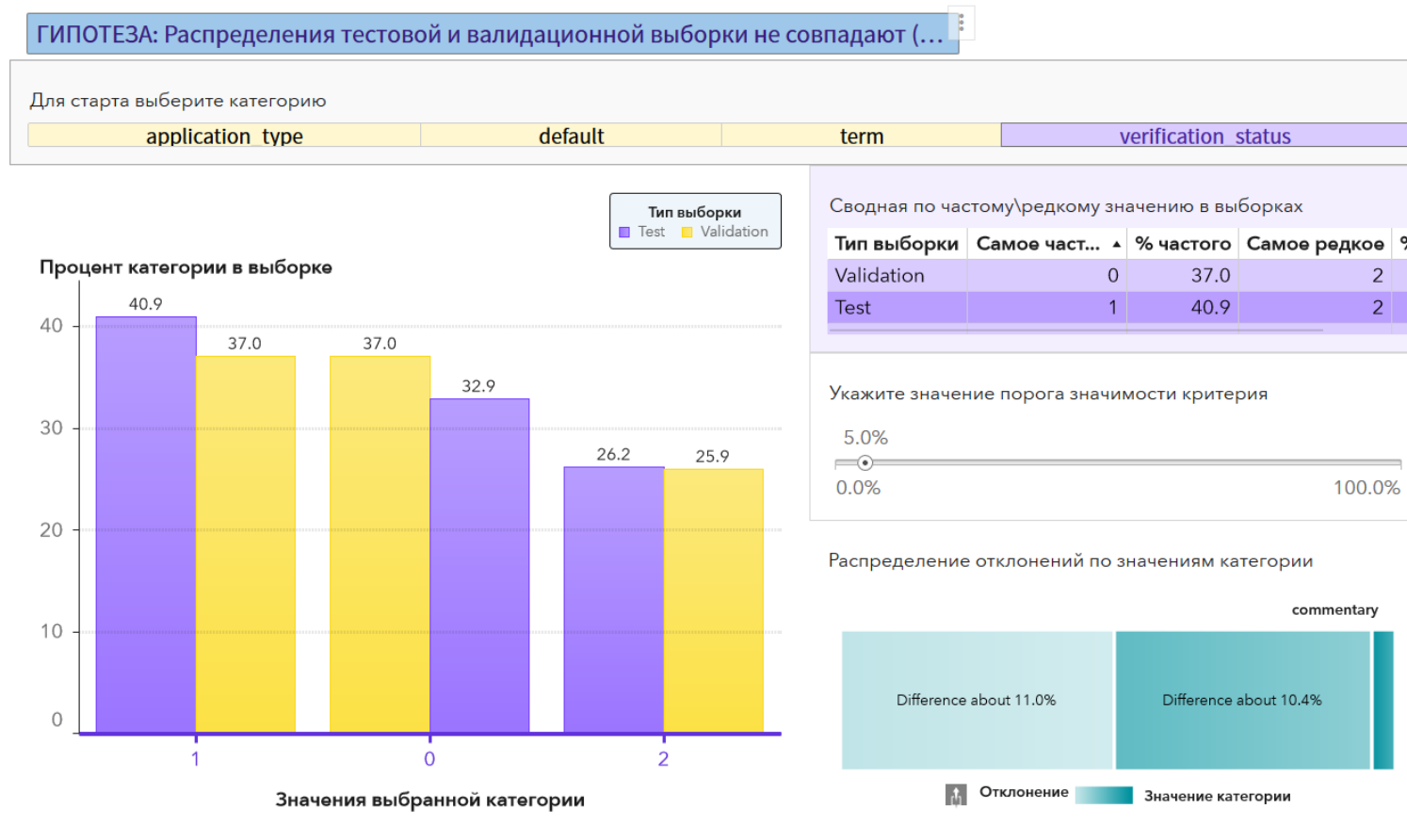
Укажите значение порога значимости критерия

2.98%

0.00%

100.00%

Пример пользовательского отчета изменение распределения категориальных переменных



Исследование Deloitte: у большинства компаний
количество реализованных проектов машинного обучения составляет

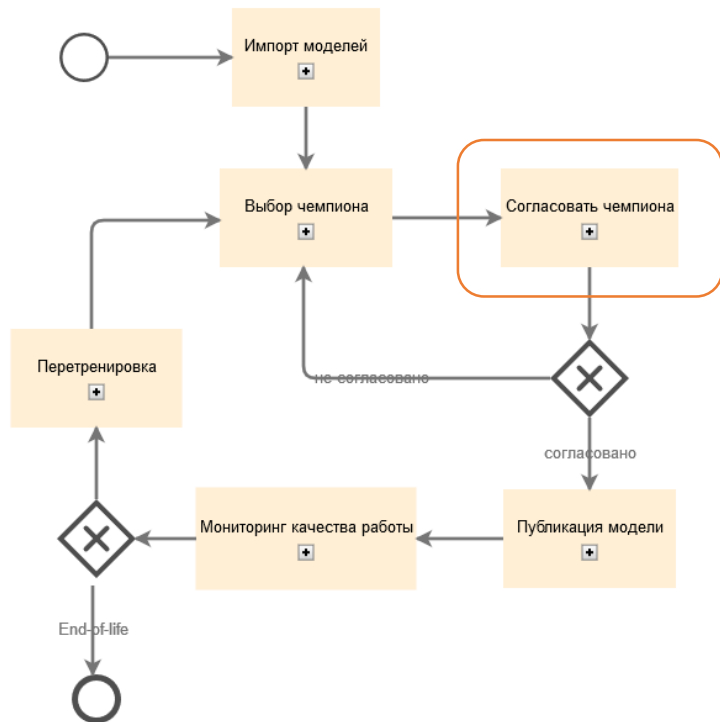
>10

+ столько же пилотных проектов *

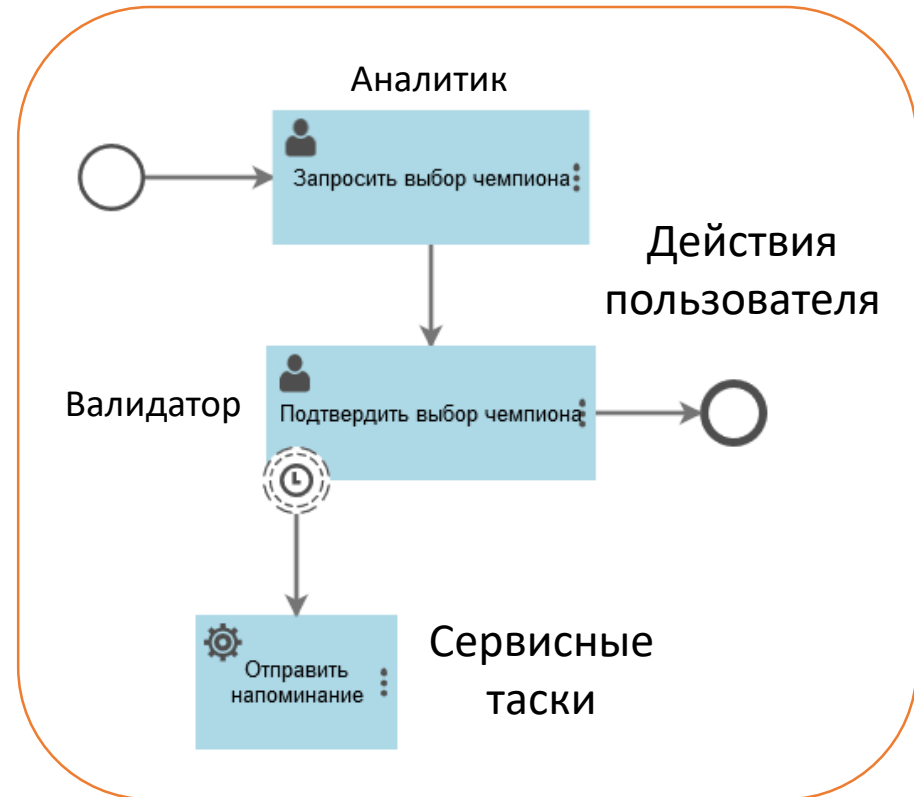
* [Deloitte: Machine learning: things are getting intense, 2018](#)

Моделирование любого бизнес процесса с помощью универсального стандарта BPMN 2.0

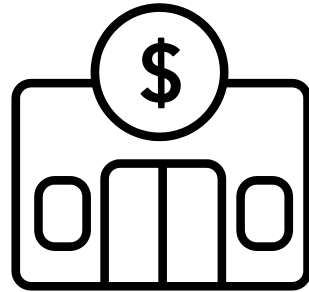
общий вид диаграммы



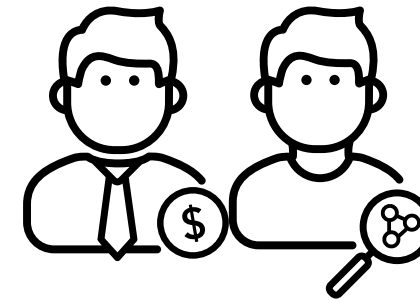
детальный вид участка диаграммы



Пример управления моделями Open Source с помощью рабочего процесса SAS Model Manager



Финансовая организация
кредитный скоринг
большое количество моделей



Роли:
аналитик
начальник отдела

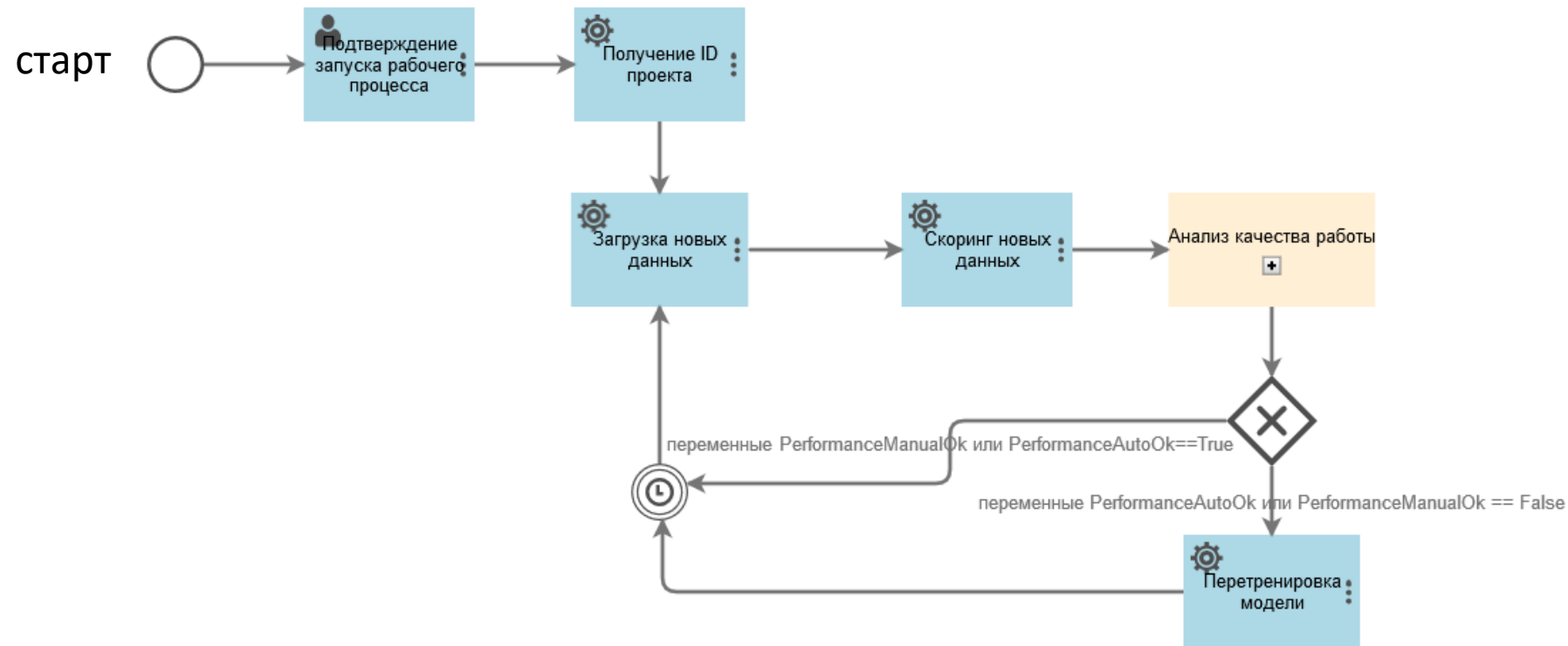
Решение:

Создание шаблона рабочего процесса, при котором

- фокусные модели отслеживаются аналитиками **вручную**
- остальные – **автоматически**, с помощью зафиксированных критериев качества

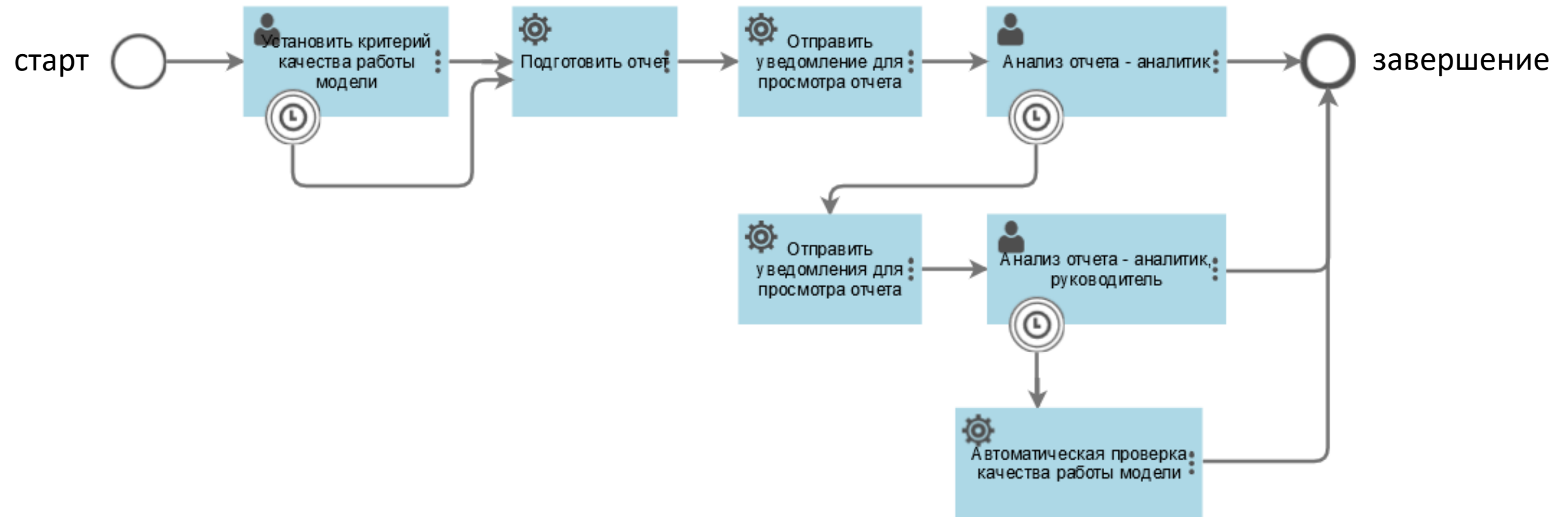
Пример управления моделями Open Source

общий вид диаграммы рабочего процесса

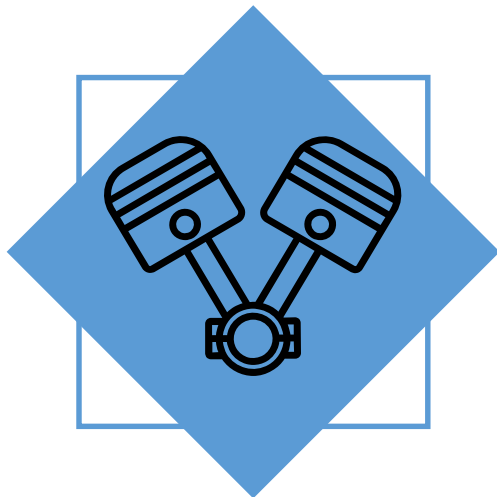


Пример управления моделями Open Source

раздел анализ качества работы

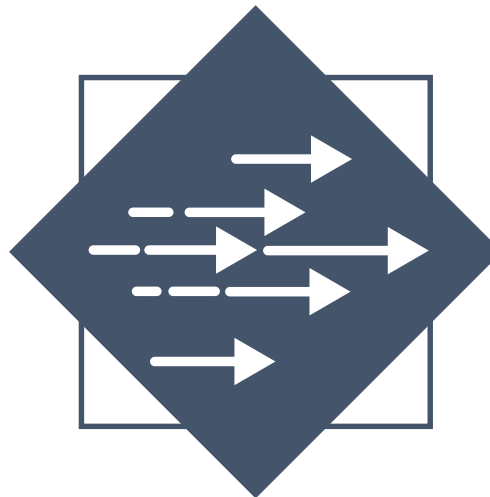


SAS Model Manager – последняя миля на пути к сквозному аналитическому процессу



Движок

Две встроенных среды
применения моделей



Оркестратор

Создание любых
рабочих процессов



Открытость

Веб интерфейс/код
SAS/Open Source

Спасибо за внимание