

СИСТЕМА РАСЧЕТА ПЛАНОВОГО СРОКА ВЫПОЛНЕНИЯ ЗАДАЧ В СЭД DIRECTUM RX

Аннотация. Оценка срока выполнения задачи играет важную роль в работе компании. Ошибочный выбор планового срока является одним из ключевых факторов, влияющих на загруженность сотрудников компании. В данной статье описана разработка системы, позволяющее рассчитать плановый срок выполнения задачи в СЭД Directum RX.

Ключевые слова: DIRECTUM RX, Случайный лес, REST, машинное обучение, регрессия.

Введение

В настоящее время основной тенденцией развития документооборота и взаимодействия сотрудников внутри компаний является использование различных информационных систем, таких как СЭД (Система электронного документооборота). Это позволяет автоматизировать бизнес-процессы компании, упростить работу с документами, заменив бумажный документооборот на электронный, а также обеспечить взаимодействие пользователей внутри системы. Однако данные системы не имеют достаточных механизмов управления загруженностью сотрудника.

Одним из ключевых факторов, влияющих на загруженность сотрудника, является ошибочный выбор планового срока выполнения поставленной работы [1].

При выставлении планового срока возникают следующие проблемы:

- сотрудник, выставляя срок выполнения задачи, не знает о загруженности исполнителя;
- плановый срок выполнения значительно отличается от реального;

- информация о схожих выполненных задачах не учитывается;
- разные инициаторы могут выставлять разные сроки для похожих задач.

Решение данных проблем позволит влиять на загруженность исполнителя и уменьшить число просроченных задач.

Существуют различные методы анализа загруженности сотрудника.

1. Метод хронометража вручную — сотрудники самостоятельно ведут записи, отмечая сколько времени потратили на выполнение тех или иных работ.
2. Метод тестирования/интервью — сотрудники заполняют специально разработанные анкеты или отвечают на ряд вопросов устно.
3. Использование программных средств, таких как CrocoTime. Данные программы отражают такую информацию, как отработанное время сотрудника, время, потраченное на выполнение определенной задачи и т.д. Это позволяет оценить нагрузку персонала, выявить факторы, мешающие эффективной работе [2].

Однако у данных методов есть несколько существенных недостатков: собранную информацию анализирует человек, что является дополнительной нагрузкой на сотрудника и требует дополнительного времени на анализ. Данные методы только собирают информацию и не могут динамически влиять на создание новых задач, т.е. использоваться сотрудниками в процессе работы, поскольку дополнительный анализ загруженности перед формированием задачи лишь увеличит время формирования задачи, поэтому для решения проблемы выставления планового срока выполнения задачи необходимо использовать иной подход.

Современные СЭД для своей работы используют методы машинного обучения, в основном с помощью этих средств решаются такие задачи, как классификация документов [4].

На основе такого подхода предлагается использовать методы машинного обучения в качестве способа для расчета срока выполнения

поставленной задачи, позволяющего прогнозировать количество рабочих часов для выполнения задачи.

Для обеспечения возможности использовать имеющуюся информацию в системе DIRECTUM RX о загруженности исполнителя, его навыках и прочем необходимо создать отдельный сервис, интегрируемый в исходную систему, использующий стиль архитектуры программного обеспечения для распределенных систем (REST) для взаимодействия с исходной системой.

описание разработанного приложения

Процесс разработки системы расчета планового срока выполнения задач заключался в:

- формировании данных для обучения модели;
- выборе модели, показавшей лучшую точность на тренировочных данных;
- разработке WEB приложения для обеспечения интеграции с DIRECTUM RX.

Исходные данные

Исходные данные содержали в себе информацию об инициаторе задачи, исполнителе, типе работ, связанной организации и вложениях в задачу.

В качестве целевой переменной использовалось количество рабочих часов, затраченных на выполнения задачи с точностью до 1 знака после запятой.

подавляющее большинство методов классификации и регрессии сформулированы в терминах евклидовых или метрических пространств, то есть подразумевают представление данных в виде вещественных векторов одинаковой размерности.

Одним из самых популярных методов преобразования категориальных признаков является dummy-кодирование. Суть метода заключается в том, что для кодируемого категориального признака создаются N новых признаков,

где N — число категорий. Каждый i -й новый признак — бинарный характеристический признак i -й категории.

Однако, у данного метода есть недостаток, связанный с большим увеличением числа бинарных признаков. Поэтому необходимо кодирование альтернативным способом — кодирование категорий некоторым интерпретируемым значением.

Так, на основе имеющейся информации, рассчитывалось среднее время выполнения по определенному виду работы по конкретной организации (например, вид работы - «Разработка», организация -«Org1»).

Таким образом, после предобработки исходных данных был получен набор данных состоящий 8171 примера с 31 признаком. Исходный набор данных был разделен на обучающую и тестовую выборку в соотношении 80% к 20%.

Модель случайного леса

Машинное обучение — это подраздел искусственного интеллекта, охватывающий множество алгоритмов, способных обучаться. Основными задачами машинного обучения являются:

- классификация — отнесение объекта к одному из классов на основании его признаков;
- регрессия — прогнозирование количественного признака объекта на основании прочих его признаков;
- кластеризация — разбиение множества объектов на группы на основании имеющихся признаков [8].

В данной работе решалась задача регрессии и в качестве модели прогнозирования была выбрана модель, построенная методом случайного леса. В качестве показателей качества моделей, были выбраны коэффициент детерминации (R^2), средняя абсолютная ошибка (MAE), среднеквадратичная ошибка (RMSE). На тестовом наборе они составляли:

1. $R^2 = 0,84$

2. MAE = 4,69 часа

3. RMSE = 8,93 часа

4. MAE% = 28%

Для разработки модели использовался язык программирования Python и его библиотека scikit-learn.

Интеграция приложения с DIRECTUM RX

Интеграция приложения в DIRECTUM RX состояла из 3 частей:

- обмен данными между DIRECTUM RX и системой расчета планового срока выполнения задач;
- периодическое обновление модели путем обучения ее на новых данных;
- развёртывание приложения.

Для обеспечения взаимодействия разрабатываемой системы расчета планового срока выполнения задач с DIRECTUM RX был разработан WEB-сервис.

В качестве инструмента разработки использовался Python фреймворк Flask, представляющий собой фреймворк для разработки WEB-приложений.

Данный WEB-сервис обрабатывает POST-запрос по указанному URL. В теле POST-запроса передаются данные для прогнозирования. В качестве ответа сервера передается спрогнозированное количество рабочих часов.

Для отправки запросов на разработанный сервис необходимо осуществить прикладную разработку на стороне DIRECTUM RX.

Прикладная разработка осуществляется в специальной среде разработчика «Development Studio». Данный инструмент позволяет разрабатывать прикладные решения на языке C# для DIRECTUM RX, с возможностью подключения сторонних библиотек, написанных на языке C#, которые позволяют расширить возможности разработки.

Прикладная разработка в DIRECTUM RX включает в себя новый вид задачи и создание специального фонового процесса для сбора данных для обучения модели.

Если в случае прикладной разработки в DIRECTUM RX данные механизмы уже реализованы, то для переноса и развёртывания сервиса предусмотрена контейнеризация созданного приложения.

Данный подход позволяет развернуть приложение изолированно от места развёртывания, сохранив все его необходимые зависимости. Это дает возможность упростить процесс переноса и обезопасит от конфликтов версий (в случае, если данное приложение будет установлено непосредственно на сервер без применения технологий виртуализации).

Для обеспечения обмена данными между сервисом и DIRECTUM RX, на стороне DIRECTUM RX был разработан особый вид задачи, в рамках которой реализован механизм отправки данных на разработанный сервис и получение результатов прогноза (рисунок 1).

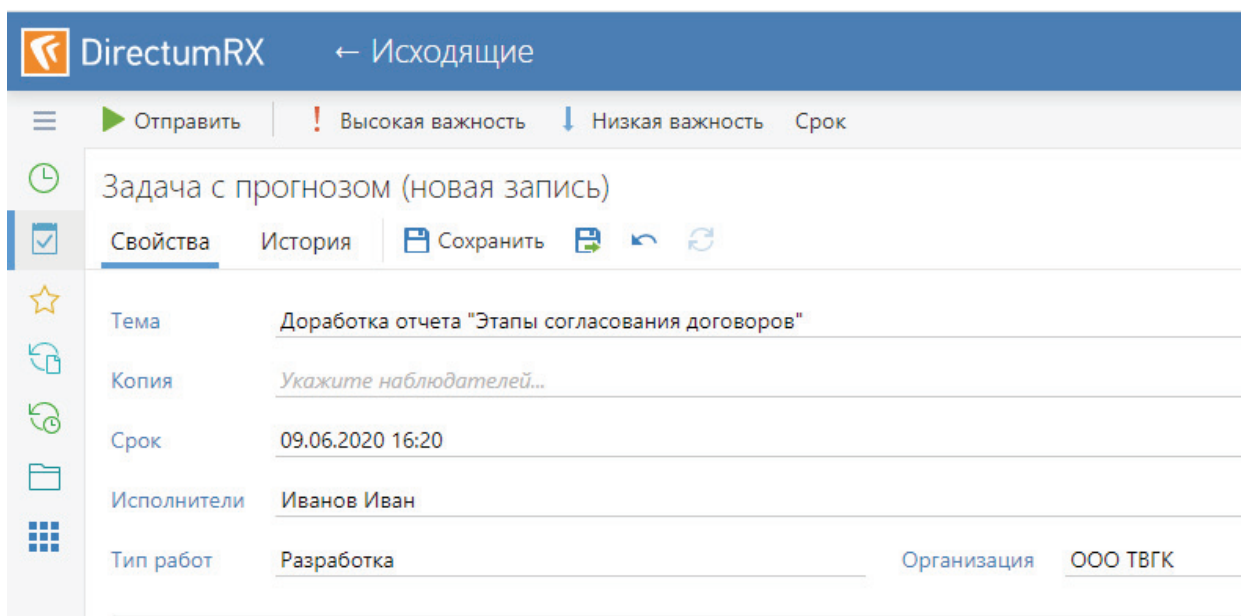


Рис. 1. Заполненная карточка разработанной задачи.

Сценарий работы предполагается следующий:

Инициатор задачи:

1. создает карточку.
2. заполняет необходимые поля.
3. нажимает кнопку «Срок», в поле «Срок» устанавливается дата и время планового срока выполнения задачи.

В случае, если прогнозируемый срок выполнения инициатора задачи не устраивает, то инициатор меняет или убирает плановый срок выполнения задачи.

4. Пользователь отправляет задачу в работу.

Для обеспечения отправки POST-запроса использовалась библиотека «RESTSharp».

Для обеспечения периодического обновления модели путем обучения на новых данных, на стороне DIRECTUM RX был разработан специальный фоновый процесс, обрабатывающий информацию, полученную из базы данных DIRECTUM RX и сохраняющий ее в .csv файл.

Так как модель используется WEB-приложением, обработка сформированного .csv файла запускается с помощью отправки запроса на сервер из фонового процесса DIRECTUM RX.

Периодичность обновления регулируется с помощью настроек фонового процесса (рисунок 2), что позволяет настроить обновление модели в определенное время и исключить попытку обновить модель в процессе ее активного использования.

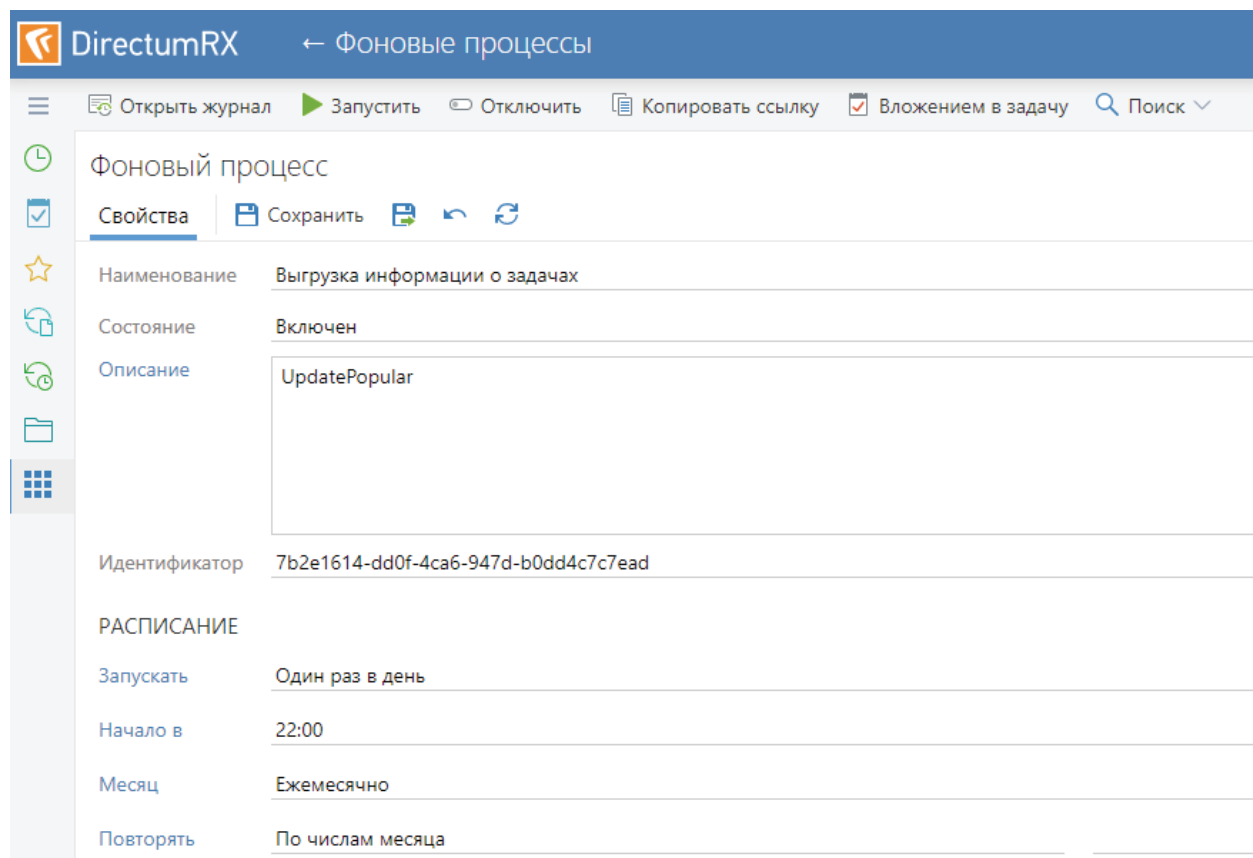


Рис. 2. Карточка разработанного фоновой процесса.

Так как созданное приложение разрабатывалось и тестировалось на отдельной виртуальной машине, а планируется использовать его в том числе и на реальных серверах, то необходимо было проработать сценарий переноса приложения с места разработки до места непосредственной эксплуатации.

Одним из самых популярных подходов к развертыванию приложений является запуск приложения в контейнере.

Контейнеризация приложения:

- позволяет запустить приложение изолировано от хост-машины;
- упрощает перенос с одного сервера на другой;
- все необходимые зависимости хранятся внутри контейнера, что и обезопасит от конфликтов версий.

Для создания и запуска контейнеров использовался Docker.

Заключение

В данной работе был проведен отбор признаков, влияющих на время выполнения задачи. Сравнение рассматриваемых моделей показало, что

лучшими характеристиками в рамках предметной области обладает модель, построенная методом случайного леса.

После сравнения моделей, позволяющих прогнозировать плановый срок выполнения задачи, был разработан WEB-сервис, обеспечивающий возможность обмена данными между DIRECTUM RX и сервисом.

В рамках интеграции между WEB-сервисом и DIRECTUM RX разработан сценарий, позволяющий обучить модель на основе новых данных и разработан специальный фоновый процесс, который собирает необходимую информацию для последующего обучения модели на новых данных.

СПИСОК ЛИТЕРАТУРЫ

1. Колесников Денис: Анализ эффективности работы сотрудников в СЭД. Данные, инструменты, визуализация. URL: <https://www.directum.ru/blog-post/918> (Дата обращения. 13.04 2020).
2. Как проанализировать загруженность сотрудников? URL: <https://crocotime.com/ru/kak-proanalizirovat-zagruzhennost-sotrudnikov/> (Дата обращения. 13.04 2020).
3. Чистяков С.П. Случайные леса: обзор. Тр. Карельского науч. центра РАН, 2013. 20 с.
4. Мандрыгина В.С. Возможности применения технологий искусственного интеллекта в задачах автоматизации электронного документооборота предприятий на примере WSSDocs. Екатеринбург: Уральский государственный экономический университет, 2019. 3 с.
5. Пальмов С.В., Денискова А.О. Случайный лес: основные особенности. Наука сегодня: теоретические и практические аспекты, 2017. 3 с.
6. Рахматова А.Ю., Косаченко А.И., Москалева А.С. [и др]. Точность методов Случайный лес и Многослойный персептрон в задаче

прогнозирования исходов детских ишемических инсультов. Уральский медицинский журнал, 2017. 5 с.

7. Андреев И. М. Описание алгоритма CART //Exponenta Pro: Математика в приложениях, 2004 48 с.
8. Открытый курс машинного обучения [Электронный ресурс]: – Электрон. текстовые дан. – Режим доступа: <https://habr.com/ru/company/ods/blog/322534/>. свободный. – Загл. с экрана (дата обращения: 06.06.2020).