

ЕГОРОВА Д. К., ГАРИН М. А.

РЕАЛИЗАЦИЯ ЗАДАЧИ КЛАСТЕРИЗАЦИИ ВОЗДУШНЫХ СУДОВ В KNIME ANALYTICS PLATFORM

Аннотация. В статье приведено создание рабочего процесса аналитической платформы KNIME Analytics Platform. Выполнена кластеризация и визуализация данных в режиме реального времени.

Ключевые слова: KNIME Analytics Platform, кластеризация, визуализация, воздушные суда.

EGOROVA D. K., GARIN M. A.

IMPLEMENTATION OF AIRCRAFT CLUSTERING PROBLEM ON KNIME ANALYTICS PLATFORM

Abstract. The creation of the KNIME Analytics Platform workflow is given in the article. Clustering and visualization of data in real time is performed.

Keywords: KNIME Analytics Platform, clustering, visualization, aircraft.

В работе приведена реализация алгоритма четкой кластеризации воздушных судов по заданным параметрам в режиме реального времени из источников с открытым API в KNIME Analytics Platform [3].

Технология ADS-B.

Будем использовать большие данные полученные с ресурса с открытым Api Flightradar24. Flightradar24 – это система отслеживающая авиасообщение всего мира в режиме реального времени. Здесь аккумулированы данные технологий ADS-B, MLAT, авиа-радаров и сведения о расписании, статусе рейсов от авиакомпаний и аэропортов [2].

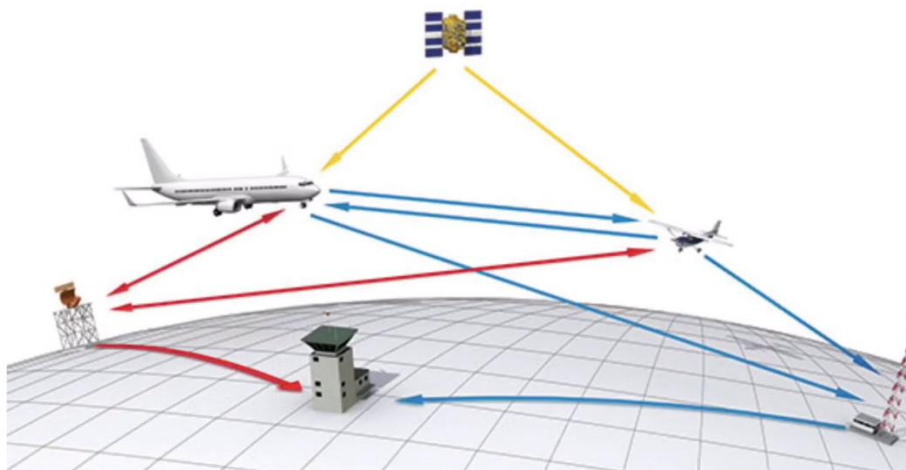


Рис.1. Схема ADS-B.

ADS-B (автоматическое зависящее наблюдение-вещание) – технология, позволяющая получать аэронавигационную информацию. Эта система использует транспондер в сочетании с GPS для передачи высокоточной информации о местоположении наземным диспетчерам, а также непосредственно на другие воздушные суда (рис. 1). Передача данных известна как ADS-B Out и ее точность выше, чем при обычном радиолокационном наблюдении. ADS-B поддерживает международный стандарт ICAO, который используется во всем мире.

ADS-B использует систему GPS для определения координат воздушного судна. Другие воздушные суда и наземные станции, оборудованные ADS-B, в радиусе порядка 150 миль принимают координаты воздушного судна, а также его тип, скорость, номер, рейс, курс, вертикальную скорость. Данная информация обновляется с интервалом времени 1 секунда по широкополосным каналам.

Организация запроса данных.

Организуем рабочий процесс в KNIME Analytics Platform для получения данных о воздушных судах в реальном времени (рис. 2). Api-запрос отправляется на Flightradar, в параметрах запроса указываются координаты двух точек, задающие область на карте в границах которой запрашивается информация.

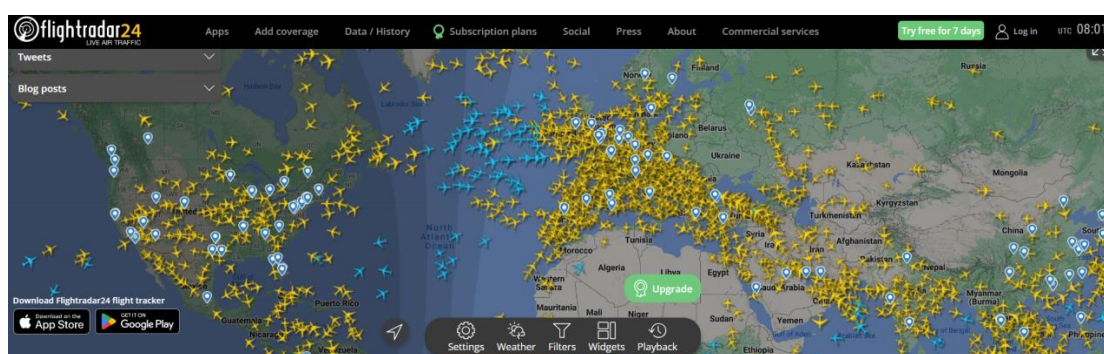


Рис. 2. Параметры запроса данных на портал.

Ресурс имеет ограничение по получению количества записей в одном api – до 1500 записей. По этой причине разобьем земной шар на 34 зоны для получения полных данных о всех воздушных судах в реальном времени.

В рабочий процесс в KNIME Analytics Platform добавим инструмент Table Creator, который позволяет создать таблицу данных. Введем параметры выделенных 34 зон в виде таблицы из 34 строк. Каждая строка включает параметры запроса с координатами северной широты и восточной долготы для заданных зон.

Используем инструмент Get Request для отправки GET запросов и получения нужных данных.

По умолчанию выходная таблица будет содержать столбец с полученными данными, типом их содержимого и кодом состояния HTTP. Узел пытается автоматически преобразовать полученные данные в тип данных KNIME на основе своего типа содержимого.

Для отправки запросов будем использовать столбец NEW URL состоящий из 34 строк. Следовательно, при запуске инструмента, последовательно будут отправляться 34 GET запроса (34 географические зоны) и записывать данные в новую колонку построчно. Инструмент добавляет еще одну колонку в таблицу с «ответными» данными в JSON формате.

В итоге, получаем порядка 1 миллиона записей, которые необходимо обработать.

Предварительная обработка данных.

Уберем лишние столбцы из таблицы с данными, оставив только столбец с JSON ответом. Для этого воспользуемся инструментом Column Filter (см. рисунок 3).

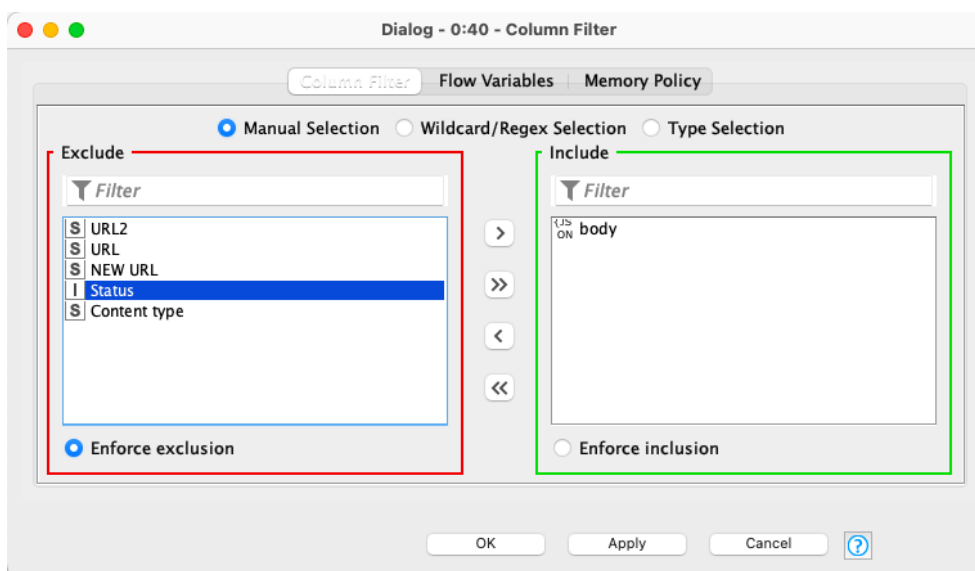


Рис. 3. Сортировка столбцов.

Рассмотрим подробнее данные которые «приходят» с сайта, для решений о дальнейшей выборке информации. Рассмотрим на примере одной из последних строк в первом из запросов.

```
1503 "2baa44c2":["155C33",57.36,20.11,196,34000,458,"F-EYSA4","SU95","RA-89139","PKV","KGD"]
```

Рис. 4. Пример массива данных одного из летательных аппаратов.

Каждая строка содержит массив данных (рис. 4):

"2baa44c2" – ключ строки в ответе, не несет смысловой нагрузки;

"155C33"– ICAO 24-bit ADDRESS (ICAO адрес ответчика с летательного аппарата);

"57.36"– широта;
 "20.11"– долгота;
 "196"– скорость в узлах;
 "3400"– высота в футах;
 "458"– курс;
 "F-EYSA4"– номер в системе;
 "SU95"– модель Самолета (Sukhoi Superjet 100);
 "RA-89139" – бортовой номер самолета;
 "PKV"– аэропорт вылета (Псков);
 "KGD"– аэропорт прилета (Калининград).

Для разбиения больших данных будем использовать инструмент JSON Path. JSON Path – это язык запросов для JSON, аналогичный XPath для XML (рис. 5).



Рис. 5. Фрагмент окна настроек узла JSON Path.

Для кластеризации данных, выберем нужные значения. ICAO ADDRESS используется для точного определения летательного аппарата и дальнейшего поиска информации о нем в системе. Так же будут взяты такие параметры как высота в футах, скорость в узлах, широта и долгота. Все численные данные, кроме адреса самолета будут типизированы как double.

При выборе 34 географических зон, для большей точности и недопущения пропуска какого-либо летательного аппарата, многие из зон были взяты с наложением друг на друга. В связи с этим возникает проблема дублирования данных. Так же во время выполнения запросов самолет может просто перелететь из одной зоны в другую что отобразится двойной записью данных для соседних зон. Для решения данной проблемы используем инструмент GroupBy. Он группирует строки таблицы по уникальным значениям в выбранных столбцах группы.

Будем искать дубликаты по двум параметрам, которые не меняются у летательных аппаратов и явно характеризуют их: это ICAO 24-BIT ADDRESS (String). Кроме этого, введем

дополнительный столбец SortColumn куда запишем бортовой номер судна (рис. 6). Дубликаты удаляются на основе этих параметров.

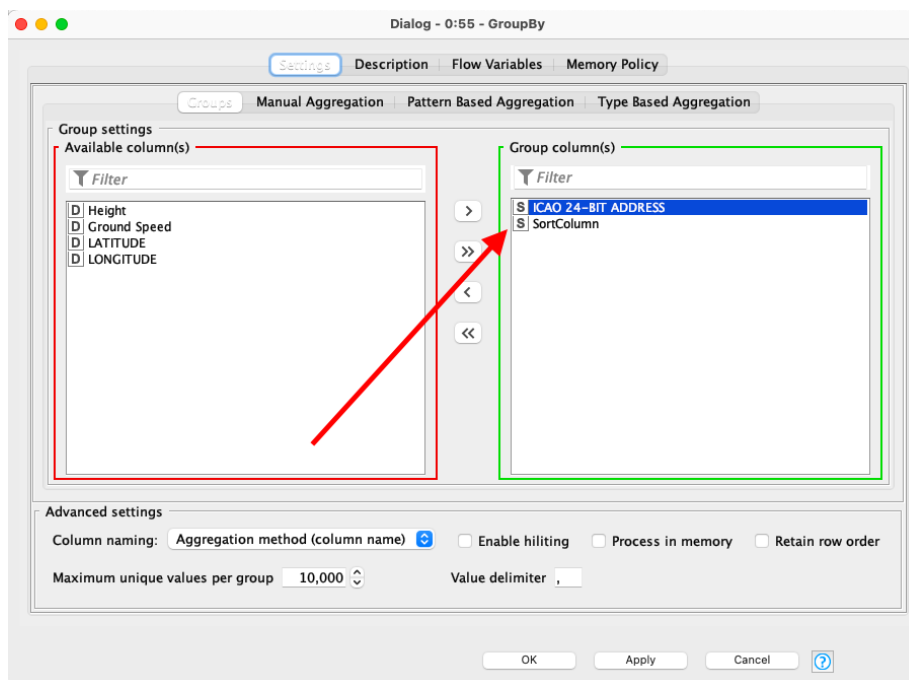


Рис. 6. Выбор критериев для поиска дублей.

Что бы визуализировать обработанную базу данных летательных аппаратов воспользуемся инструментом OSM Map View. Он не входит в стандартный пакет при установке KNIME. Загрузим его отдельно с открытого репозитория инструментов KNIME Hub, раздел OsmMapViewNodeFactory. Он предоставляет интерактивный вид карты мира. При желании на него также могут быть нанесены некоторые маркеры по данным широты и долготы.

На рисунке 7 по данным обработанной таблицы визуализированы все 13302 летательных аппарата (данные от 17 мая 2022 года).

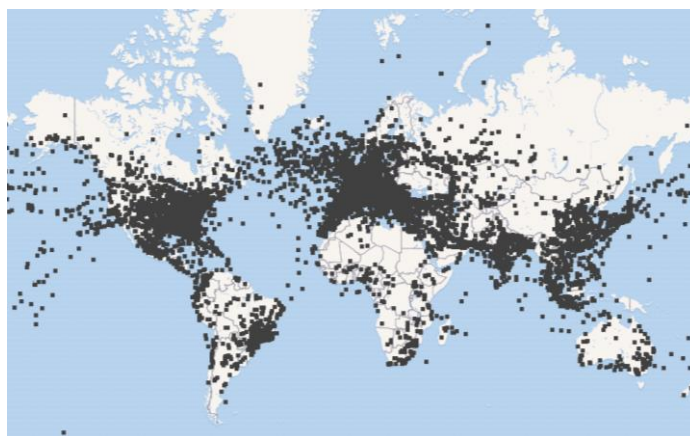


Рис. 7. Все летательные аппараты на карте мира.

Кластеризация летательных аппаратов по скорости и высоте.

Решим задачу кластеризации по двум параметрам: скорости и высоте. С помощью инструмента Column Filter были взяты из колонок таблицы, которые непосредственно нужны для данного примера кластеризации.

Выполним настройку узла кластеризации k-Means (рис. 8) исходя из проверки эффективности кластеризации методом силуэтов (инструмент Silhouette Coefficient) [1].

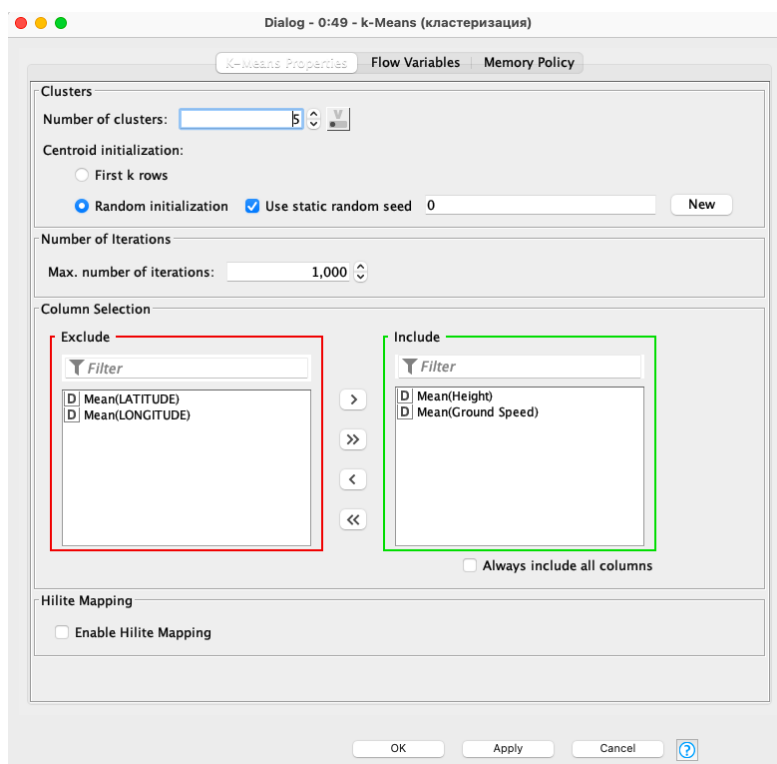


Рис. 8. Меню настроек k-Means.

File Edit Hilite Navigation		
Row ID	D	Mean ...
cluster_1		0.75
cluster_2		0.611
cluster_3		0.448
cluster_0		0.522
cluster_4		0.509
Overall		0.575

Рис. 9. Коэффициенты силуэтов.

Общее значение коэффициента силуэтов равно 0.575 (рис. 9).

Отсортируем таблицу по возрастанию кластеров и с помощью инструмента Color Manager окрасим точки каждого кластера в разные цвета.

Для вывода данных используем узел для построения точечных графиков Scatter Plot (рис. 10). Это точечная диаграмма с использованием библиотеки диаграмм на основе JavaScript. Доступ к представлению можно получить либо через действие «интерактивное представление» на исполняемом узле, либо на странице веб-портала KNIME Server.

По оси абсцисс отмечена скорость в узлах, по оси ординат высота в футах.



Рис. 10. Распределение данных по кластерам.

Интерпретация результатов

Проведем соответствие каждого кластера техническим данным полетов, классу и состоянию воздушного судна.

Рассмотрим кластер 1 (cluster_1, красный цвет). Здесь наблюдаем относительно небольшую скорость и высоту летательных аппаратов. Так же есть скопление точек около нулевых отметок. Это самолеты, которые совершили посадку и стоят на земле с нулевой высотой и скоростью. Точки, которые имеют нулевую высоту, но уже возросшую скорость, от 20 до 150 узлов можно соотнести с самолетами, которые совершают руление по аэродрому, а также идут на взлет по полосе. Это подтверждается техническим данным – скорость руления 30 узлов, скорость взлета среднемагистральных самолетов (Boeing 737) 145 узлов. Точки, имеющие небольшую высоту и скорость соответствуют самолетам, которые взлетели и набирают крейсерскую высоту или идут на посадку. Таким образом, кластеру 1 (cluster_1, красный цвет) соответствуют самолеты, которые находятся на земле, взлетающие или совершающие посадку.

Рассмотрим кластер 2 (cluster_4). Здесь разброс высот составляет примерно от 6 тыс. до 15 тыс. футов. А скорость варьируется от 60 до 290 узлов. Этому кластеру можно отнести

вертолёты, высота полета которые соответствует данным, а скорость невелика. Так же к этому кластеру отнесем легкомоторные самолеты, такие как АН-2 (кукурузник) или Cessna 172 (самый популярный легкомоторный самолет в мире). Их параметры полета по высоте и скорости соответствуют данным значениям.

К кластеру 3 (cluster_0, зеленый цвет) соответствуют высоты от 20 тыс. до 30 тыс. футов. В этот кластер можем отнести ближнемагистральные и среднемагистральные самолеты гражданской авиации.

К кластеру 4 и 5 (cluster_3, cluster_2) будут соответствовать воздушные суда, проходящие на больших эшелонах. Это дальнемагистральные и межконтинентальные пассажирские и грузовые самолеты.

Также наблюдаем несколько точек выброса. Это такие летательные аппараты, которые имеют нулевую скорость и большую высоту или очень низкую высоту и запредельную скорость. Скорее всего такие данные свидетельствуют о неисправности оборудования или возникновении аварийных ситуаций. Все исследованные точки выброса на данных от 17 мая 2022 года 15:00 МСК, оказались технической неисправностью оборудования со стороны приемника. Также в точку выбросов, предположительно, попали два военных истребителя, которые работали на малых высотах с большой скоростью.

Для наглядности, с помощью инструмента OSM Map View, были выведены на карту мира все летательные аппараты с цветами точек соответствующих кластерам (рис. 11).

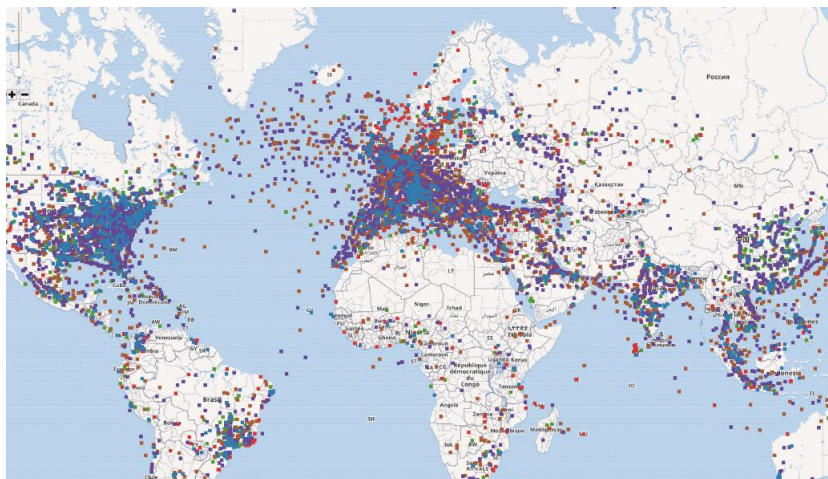


Рис. 11. Летательные аппараты с цветами точек по кластерам.

Также на карту выведем кластер 1 – самолеты, которые находятся на земле, взлетающие или совершающие посадку. Скобление таких самолетов происходит вблизи больших городов и аэропортов (рис. 12).



Рис. 12. Самолеты, которые находятся на земле, взлетающие или совершающие посадку.

При выводе на карту кластера 2 – это воздушные судна, проходящие на больших эшелонах, дальнемагистральные и межконтинентальные самолеты. Можем наблюдать линии высотных воздушных трасс (рис. 13).

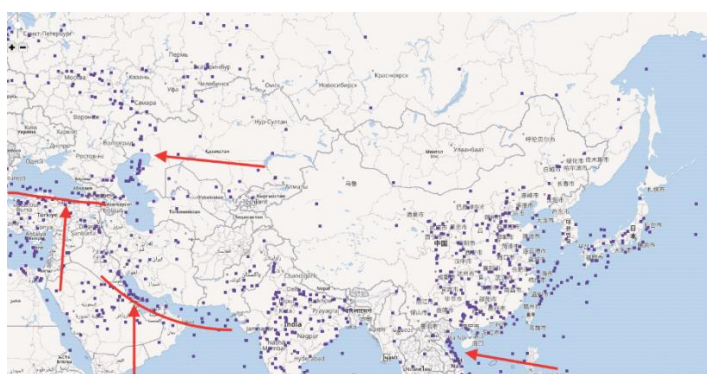


Рис.13. Линии воздушных трасс.

Таким образом, в KNIME Analytics Platform был построен рабочий процесс, который может представлять практическую ценность при решении задач визуализации воздушных судов в режиме реального времени.

СПИСОК ЛИТЕРАТУРЫ

1. Гарин М. А., Егорова Д. К., Сайфетдинов С. Ф. Создание workflow аналитической платформы KNIME для анализа данных на примере вакансий сайта HeadHunter [Электронный ресурс] // Огарев-online. Раздел «Физико-математические науки». – 2021. – №12. – Режим доступа: <http://journal.mrsu.ru/arts/sozdanie-workflow-analiticheskoy-platformy-knime-dlya-analiza-dannyx-na-primere-vakansij-sajta-headhunter> (дата обращения: 20.11.2021).
2. Flightradar24: Live Flight Tracker Map [Электронный ресурс]. – Режим доступа: <https://www.flightradar24.com/54.18.45.18/6> (дата обращения: 15.05.2022).
3. KNIME Analytics Platform Map [Электронный ресурс]. – Режим доступа: <https://www.KNIME.com/KNIME-analytics-platform> (дата обращения 10.05.2021).