Боброва Е.В., Петрова А.В.

*Process Mining*

**Отчет по лабораторной работе №2**

*О журнале событий*

Процесс, рассматриваемый в BPI Challenge 2018, охватывает обработку заявок на прямые выплаты ЕС для немецких фермеров из Европейского сельскохозяйственного гарантийного фонда.

Изначальный файл с данными был очень большой и коллаб плохо справлялся с его открытием, поэтому для анализа мы взяли его фрагмент. Для работы с журналом событий определим основные атрибуты событий в журнале (Табл. 1).

Таблица 1 – Основные атрибуты событий в журнале

|  |  |
| --- | --- |
| org:resource | Ответственный департамент |
| activity | Выполненная деятельность, на завершение которой указывает данное событие. |
| concept:name | Уникальный идентификатор обращения для приложения |
| application | Идентификационный номер заявителя, одинаковый в разные годы |
| selected\_risk | Было ли приложение выбрано для проверки из-за оценки риска? |
| payment\_actual{x}\* | Сумма (в евро), фактически полученная заявителем. О значении {x} см. выше. |
| time:timestamp | Время, когда произошло событие. Обратите внимание, что порядок событий с одинаковыми временными метками нельзя вывести из файла. Также обратите внимание, что некоторые метки времени вводятся вручную и поэтому могут содержать орфографические ошибки. |

К остальным атрибутам относятся статус платежа, типы документов, заметки, штрафы и тд.

*Статистики по журналу событий*

При обработке данных журнала была получена следующая информация:

Количество уникальных заявителей 4771.

При первичном просмотре данных могло показаться, что колонки case:young farmer и case:small farmer одинаковые, оказалось что нет.

Однако case:concept:name и case:application одинаковые и одну из них можно удалить.

Всего в кейсе 4058 различных вариантов путей. Причем самый популярный путь повторяется только 87 раз, что говорит о довольно уникальных путях.

Суммарно 289 лет было затрачено одним департаментом на решение.

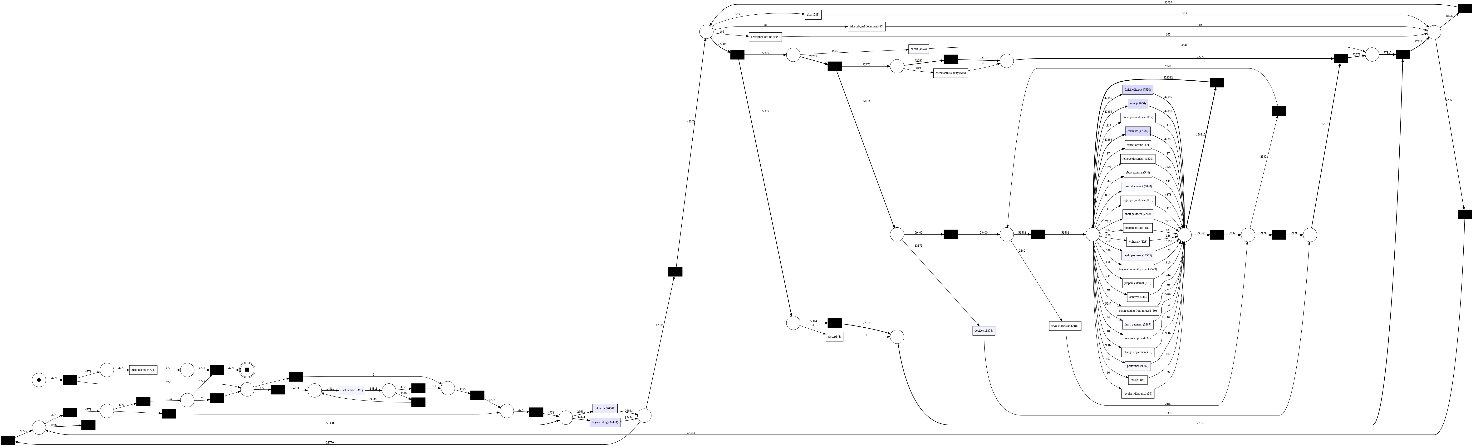
*Построение моделей*

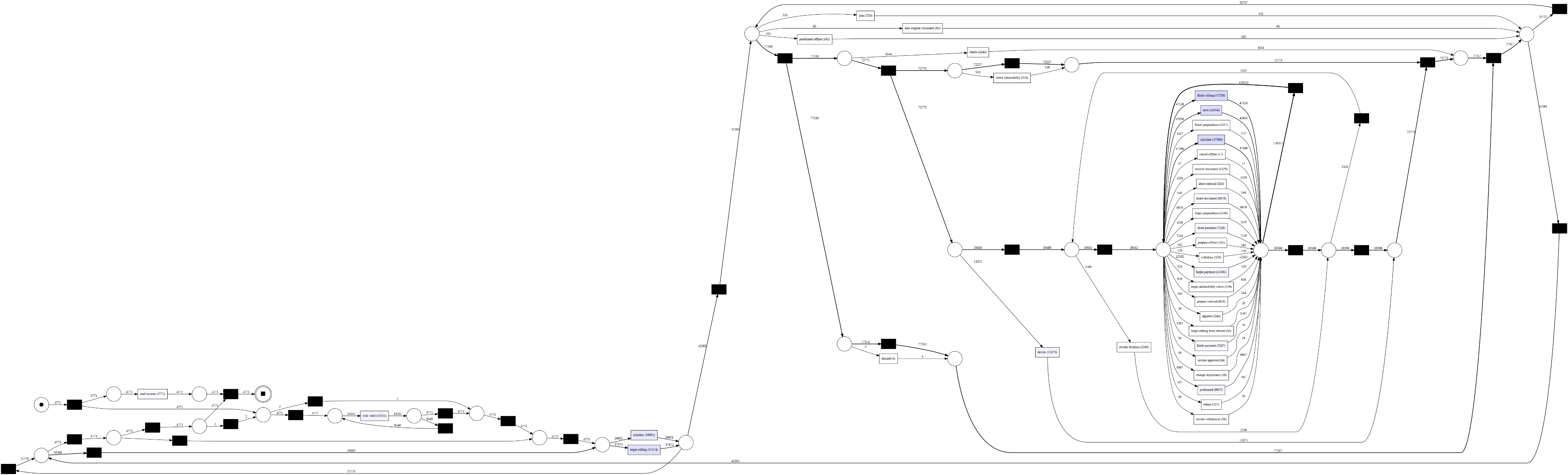
Была построена сеть петри, с помощью алгоритма alpha miner, для наблюдения всех возможных событий.

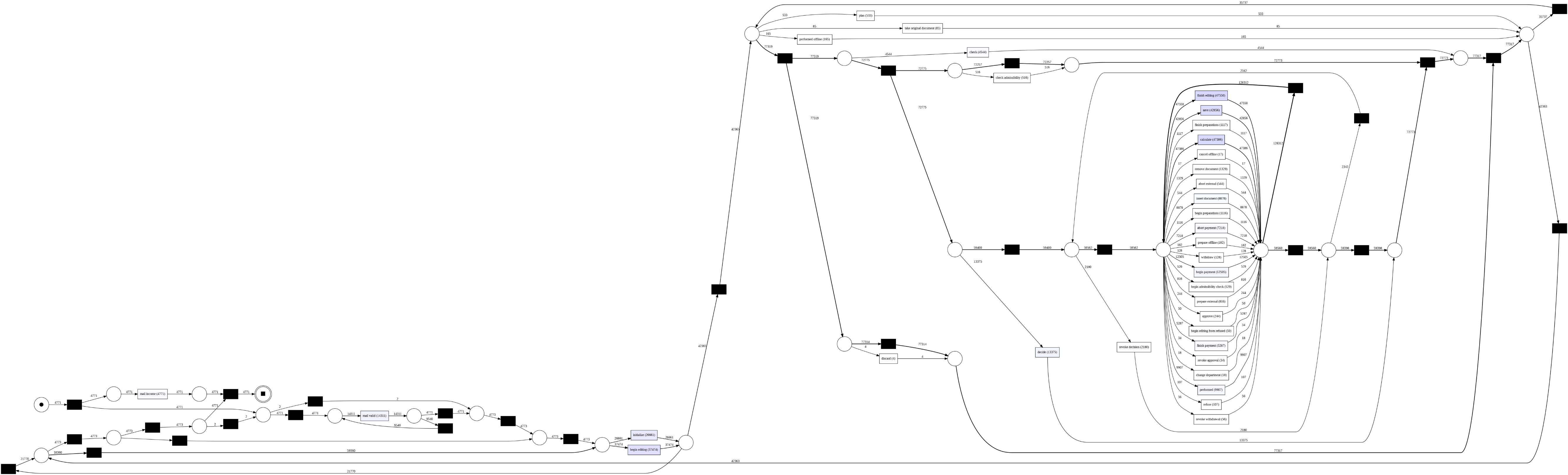


Вид диаграммы можно объяснить тем, что alpha miner не обрабатывает петли.

Второй метод показал себя гораздо лучше. А именно Inductive miner.







Простота равна 0.5227272727272727

Пригодность воспроизведения, используя метод на основе токенов, равна {'perc\_fit\_traces': 99.95808006707189, 'average\_trace\_fitness': 0.999999419420082, 'log\_fitness': 0.9999973680262566, 'percentage\_of\_fitting\_traces': 99.95808006707189}

Пригодность воспроизведения, используя метод выравнивания, равна {'percFitTraces': 100.0, 'averageFitness': 1.0, 'percentage\_of\_fitting\_traces': 100.0, 'average\_trace\_fitness': 1.0, 'log\_fitness': 0.9996464696123124}

Точность с помощью метода ETConformance равна 0.1315827319299122

Для heuristics miner граф получился слишком большой из-за чего сложно оценить его составляющие.

