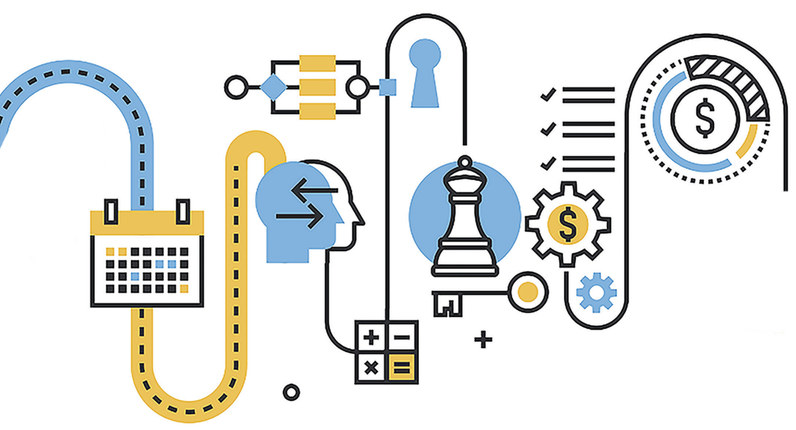


Italian Road Fines System:
A process mining application

Introduction

Who has not experienced the situation that you submitted a request into some administrative system? Administrative processes may take longer than expected, which subsequently may cause confusion and frustration among people. A tool to estimate the time it takes before the next step in an administrative process initiates could provide people with clarifying insights into the time such a process will take.

Dataset

The data set that is used to train the prediction model is a data set containing **road traffic fine management processes in Italy**. The data set contains information about distinct cases of road fine processes, where each distinct case consists of a sequence of events.

150,370 cases and **561,470 events.**

The data set is divided into two parts: **a training set and a test set, using a 80/20 percent split**.

The data set contains the **timestamp** at which the events happened.

Our tool also works on 3 additional datasets: **BPI Challenge 2017** & **BPI Challenge 2018** & **BPI Challenge 2012**

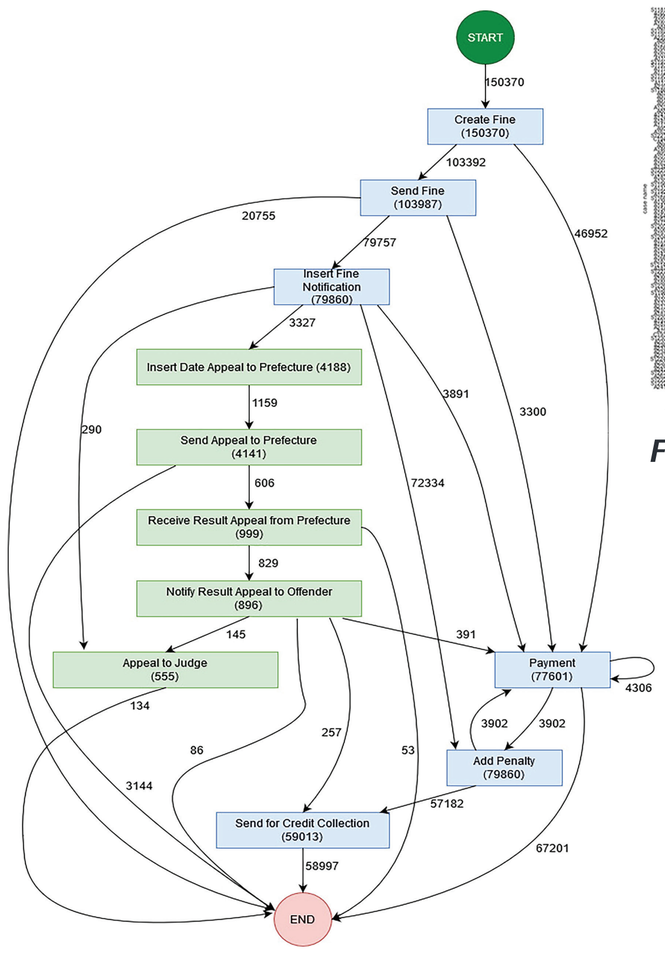


Fig. 1 *Flowchart of the proces in our dataset*

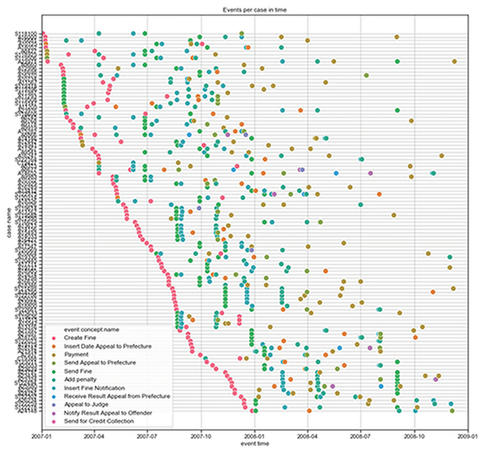


Fig. 2 Scatter plot of the events grouped by case

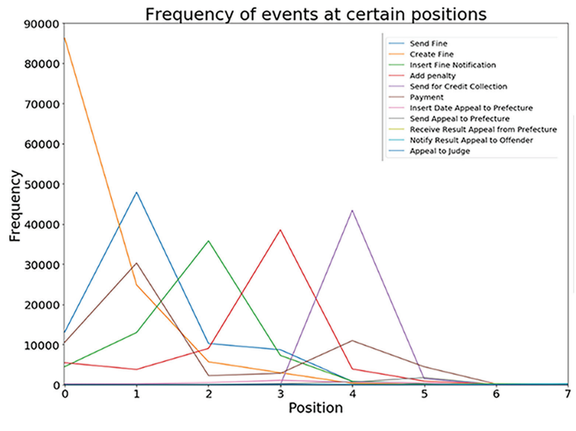


Fig. 3 Frequencies of events on positions

Baseline Model

Our baseline model predicts the **next activity** that will happen in the road fine management process, and **how much time** activities will take in the process.

Given a specific step in a case, the baseline model will predict the **most frequently occurring event at that step after** as the next event in that case.

The prediction of how much time an event will take is done by computing **the average time it takes to proceed to the next event**, for every step in the process.

Prediction Model

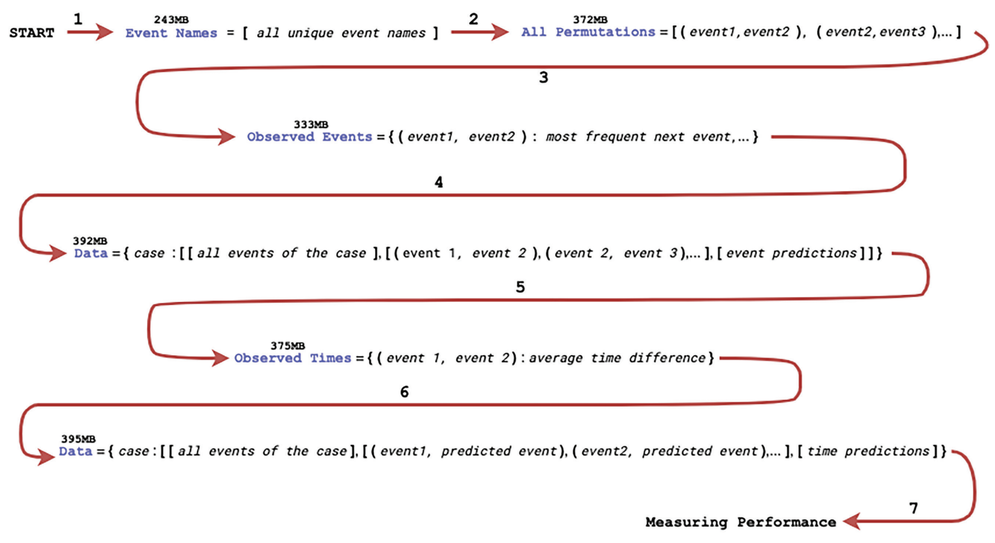
Version 1

Feature Selection

The day of the week of the last event in the permutation is a selected feature.

Explainability

The model works based on the following steps:



Version 2

Another prediction model was created to see whether the performance of the second implemented algorithm could be improved over.

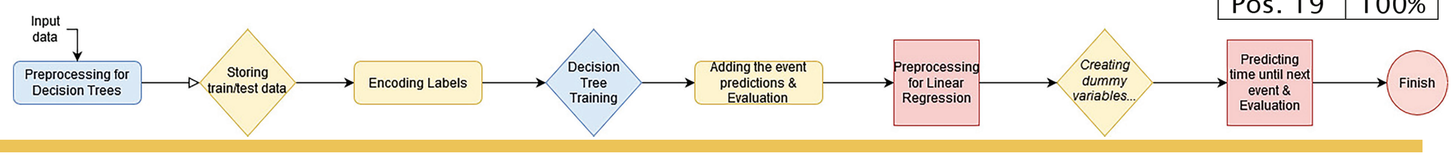
Feature Selection

The day of the week of event in a case is also a selected feature for the time prediction model.

Explainability

The model works based on the following steps:

1. **Decision trees** are used to predict the next event in a case.
2. Multiple decision trees are created to predict which event will occur at each possible step in a case, based on the full trace of events in the case.
3. **Linear regression** is used to predict the time differences in between events.
4. The regression model predicts based on the type of event, the position of the event in the case, the previous event in the case, the day of the week the previous event occurs, and the number of events in the case.



Accuracy of next event prediction per position (each decision tree):

Pos. 0	58%
Pos. 1	75%
Pos. 2	91%
Pos. 3	44%
Pos. 4	87%
Pos. 5	84%
Pos. 6	40%
Pos. 7	34%
Pos. 8	79%
Pos. 9	68%
Pos. 10	33%
Pos. 11	100%
Pos. 12	67%
Pos. 13	67%
Pos. 14	0%
Pos. 15	0%
Pos. 16	0%
Pos. 17	0%
Pos. 18	0%
Pos. 19	100%

Discussion

Error

	Italian Road Fines Data		BPI Challenge 2017		BPI Challenge 2018		BPI Challenge 2012	
Tool	**Accuracy**	**RMSE**	**Accuracy**	**RMSE**	**Accuracy**	**RMSE**	**Accuracy**	**RMSE**
Baseline	22%	~137.7 days	3%	~5.3 days	21%	~16.6 days	9%	~16.5 days
Version 1	60%	~134.4 days	68%	~5.8 days	47%	~16.4 days	72%	~16.6 days
Version 2	69%	~166.8 days	84%	~15.5 days	58%	~16.5 days	79%	~15.8 days

*As reference, the average time between consecutive events is ~115 days in the Italian Road Fines. Accuracy and Root Mean Squared Error are common evaluation measurements suitable for our tool.

Performance

The current tool runs in command line and it takes ~ 300 seconds to output the result. It never takes more than 430MB of RAM. This performance is achieved on a computer with Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80GHz and 8GB installed RAM and recorded using the memory_profiler package.

Conclusion

Initially, we designed our algorithm to make predictions based on the frequency of events in the current position. We improved over this algorithm by predicting the next event in a case based on permutations of every 2 events within a case. We also managed to implement the full trace of events for the time prediction for this model. In addition, a third prediction tool which uses decision trees and linear regression was implemented. This third model was created using online techniques. Lastly, we managed to implement the algorithms on three data sets. So, all in all, we created a prediction tool that provides people with clarifying insights into administrative processes, as desired.

2IOI0 DBL Process Mining Group 13

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

TILBURG UNIVERSITY

Jovan Saković 1399365
Stefan Stoev 1397664
Gijs Toemen 1397672
Iulia Trandafir 1399381
Kaloyan Videlov 1398091
Wouter van der Wee 1256092