Project ID: 99

Project Title

Infomedia Process Miner

Client Name

Kim Horn

Group Capacity

3 groups

Project Background

The goal is to create an application to support the mining, understanding and improvement of automotive business processes and tasks from transactional event data. Over the last 8 years Process Mining has matured from an obscure Machine Learning method to a mature industry wide Data Mining approach, supported by the big 5 consultancies, and many commercial products. The approach was invented by William van Der Aalst, at Eindhoven University, and published in his book "Process Mining – Data Science in Action", Springer. See https://www.processmining.org

https://en.wikipedia.org/wiki/Process_mining

The main objective of Process Mining is to offer fact-based insights that come from actual data, as people execute a process daily. As these people execute their business processes, in an organization, they leave behind a trail of data. This Event Data can be used to extract process-related information, such as automatically discover the actual process model, or check for undue costs, delays or gaps. Management may have a perception of what their processes are, but often the reality is very different, and human biases often cloud this perception from the reality. Understanding the reality, by visualizing a process, provides the first step to quality and cost improvements. Process Mining also helps you audit and analyze the processes as they run. Process mining replaces, the old attempts at BPM, that was more manual, more tedious, and more subjective. It's speedier and cheaper, and its objectivity increases its accuracy.

In the automotive industry, dealerships have many processes, including after-sales, customer support, service and repair, accident repair, warranty claims, non-warranty work, parts purchasing, order handling, stock procurement and inventory management. These processes may involve interactions with external organisations such as OEMs, parts suppliers, specialist providers, after-market suppliers, insurance companies, and payment channels.

Dealership want to reduce the cost and improve the quality of these processes, to meet or

better their KPIs. They may ask questions like, is the process these people are actually carrying out, aligned with expectations, is it performant, optimal, can it be improved, can we reduce costs. Are there delays, blockages, redundant steps, do customers wait unnecessarily, or do staff wait for other parties to complete work? Are parts available on time? Can the warranty claim processing be streamlined? What's needed to keep technicians operating at 110% efficiency? What steps can we improve to keep productivity at 85%. Are there loops or cycles in the process, e.g. rework where issues are repeatedly worked on but not fixed. Tools like ABC can be used to understand costs. There are many ways work can be improved to flow better with less friction or uncertainty.

Infomedia would like to integrate Process Mining into its product suite to help dealerships and so using a commercial product ultimately not feasible. We also do not want a generic and complex mining tool but one specific to automotive dealerships, that is simple and easy to use.

Project Scope

It is expected that the project 'way of working' will be Agile and Iterative. As such, the requirements in this document are provided to start a conversation and will be iteratively fleshed out as the project progresses.

The aim of this work is to produce a Prototype, that can be used to demonstrate and explore the idea with the business internally and our customers, to gain feedback. The term 'Prototype' is used, as this will be more than a POC, and less than an MVP. It is more than a POC, as technically process mining is proven, and not meant to be a final Infomedia product, so not an MVP.

The Prototype system may be comprised of 6 main conceptual components:

1. Process mining server - does the data management and mining. Stores Event logs and models. Contains the core mining algorithm (see 6).
2. Web User Interface to Analyse - Display and Editing front end, Chat bot approach may be beneficial to interact and question the user.
3. Web User Interface to Run – allows user to engage the process and see what to do.
4. Task miner - a component that helps capture non-official process tasks. These may occur outside the system creating the event logs, or be used to capture event data with no system. It could be a chatbot, that allows the user to tell it the status of tasks and then record them for analysis.
5. Data Generator - a tool to help generate example cases for demonstration.
6. Inductive mining algorithm – implement an IM algorithm that produces Process tress. This is to be plug and play with Open-Source code used in (1).

Potential Open-Source tools are available in Python and R:

https://bupaverse.github.io/docs/
https://github.com/bupaverse
https://rpubs.com/liam/BupaR1
https://www.r-bloggers.com/2019/03/process-mining-part-1-3-introduction-to-bupar-package/#google_vignette
https://github.com/TheWoops/awesome-processmining
https://github.com/process-intelligence-solutions/pm4py
https://github.com/pmlab/pmlab-full
https://github.com/rapidprom

ProM is an extensive Process Mining tool for Research use only. It can be used as a gold standard to compare models with: https://promtools.org/prom-6-14/

There are many Process Mining Algorithms, e.g. alpha, heuristic mining, 'Inductive' miner. Similarly, the Process model representation approach is varied, from Inductive Trees to BPMN models, Petri-Nets, EPCs, WF-Nets, YAWL etc. A review of approaches and the feasibility of implementing them is required. Algorithms balance 4 quality dimensions: fitness, simplicity, generalization and precision. They need to cater for noisy data, scale, and be able to complete. It is known that alpha cannot manage these tradeoffs, and is not guaranteed to complete, Inductive approaches are reported as sound, and scalable. Decisions will need to be made as to which algorithms to use or implement.

It is suggested that a twofold approach is taken. One set of teams develops the user facing elements using Open -Source tools. This will deliver a working prototype with less risk. However, any constraints on Infomedia using these Open-Source components needs to be reviewed at the very start of the project and evaluated around copyright and licensing. The assumption is that they can't be used for final delivery but can be used by UNSW for this Prototype. A second team(s) is employed to mitigate this issue and develop a core algorithm. The most sound now is those based on Inductive Mining. This approach allows RDR approaches to extension to be applied.

The project could be delivered in Java, Python or Shiny. The project could be delivered as a Shiny Application, using Python or R. It could use, for example, the BupaR Open-Source package that is wrapped by a Shiny or Python Web Application.

Project Requirements

Use Case 1: Discovery - The user loads an event log file, stores the events, and starts process mining. The system automatically generates a process model and display it graphically to the user. The system displays the process basic frequencies and statistics. The user should be able to see the events that are not captured in the model and why.

Use Case 2: Conformance and audit - The user loads a new event log, and a past process model, and can run it through the model. The user can check the alignment between the observed and expected workflows. By comparing new data with a current process models they can identity differences and diagnose deviations or inefficiencies. KPIs can be used as the metrics to determine targets. This can be done in real time to address issues quickly. It can be used on past data for audit. The KPI performance gaps, impacts and issue are displayed. Includes capabilities to check conformance and compliance, not only graphically through overlays, but also through data analysis and performing gap analysis.

Use Case 3: Enhancement - Allow the user to edit and enhance the process model graphically and save different versions. The use can use the event data to improve and optimise existing processes, to improve maturity, and to meet KPI targets. The processes may be repaired or extended. The user should also be able to create a company hierarchy or matrix, showing how the actors (e.g. staff) involved in the process are organized. Depending on the purpose and available information, techniques can focus on various aspects such as flow, time, organisational structure, or costs. This helps the organisation produce more prescriptive processes. The edits may allow the user to include events not included in the generated mode, in Use Case 1. The system provides intelligent support for enhancing or extending existing or apriori process models by using additional data from recorded logs and events.

Use Case 4: Operation Support - Users engage with system to get real time support on the process. The user can update the status of the process and save the current state. The models can be used in real time to guide executing processes and tasks. An interactive User Interface will allow users to see the status of the process and task. This could be guided by an interactive chatbot.

Use Case 5: Generate - The developer users can generate new test cases to explore and test how the Prototype functions, for QA purposes. A tool is provided to generate transaction logs given a list of activities and tasks, with their probabilities. These events can be edited manually after to add in edge cases and remove unwanted ones.

Non Functional Requirements
NFR 1: provide an application with a simple and great User Experience. E.g. interacting with the process models should work without difficulty. We do not want to provide a full-on process mining suite but one that is easy to use for non-technical users.
NFR 2: provide the potential for the system to scale but initially cope with a few concurrent users.
NFR 3: allow potential white labelling and customization to align the product with Infomedia Logos for demonstration only.

At this stage other NFRs around security, interoperability, performance etc. can be traded off.

Domain Model Requirements
Data includes a set of cases, an ordered set of events; each event may have these minimum attributes:

- CaseID
- ActivityID
- Activity
- Timestamp

Other attributes of data could be:

- resource (actors - people or devices, staff, managers doing or deciding in a task
- duration
- cost
- success
- issues raised/solved
- subtasks (not tracked)

Additional optional capabilities

- GenAI capabilities to improve data preprocessing and discovery; supporting adaptive process modifications through real-time monitoring and refined recommendations; and enhancing explainability and understandability, thus broadening process mining accessibility to a wider audience.
- Advanced process

Required Skills

Understanding of Data Mining and Machine Learning, and process modelling approaches. Understanding of Process Mining algorithms.

HTML, Python, Java and Shiny application development. If Java is used, then Spring MVC and Spring Boot are preferred. Python libraries like Flask, Pandas, Django, Anvil, Reflex, NiceGui.

Generative AI, and common Python GenAI libraries, e.g. Langchain.

Expected Outcomes

The outcomes are:

- A working Prototype application covering the use cases above, with all the code, and build/run instructions.
- The full source code of the system. All build and test instructions and scripts.
- Review and document process mining algorithms, and process model representations, produce a recommendation for use.
- Review the Open-source tools on the market and choose one that meets the goals. Understand the licensing issues.

- More detailed requirements, as they are fleshed out.
- An implementation of an Inductive Mining algorithm.
- Documentation and architecture of the system with design choices explained.
- User guides.
- Research papers that were used in the project.

Disciplines

Software Development;Web Application Development;Computer Science and Algorithms;Artificial Intelligence (Machine/Deep Learning, NLP);Generative AI (GenAI);Big data Analytics and Visualization;Human Computer Interaction (HCI);

Other Resources

Infomedia will provide background information, expert knowledge, in both the domain and Process Mining ML approaches, automotive, domain and background knowledge, will be provided, with demonstrations of the existing SaaS products. We can share many references and papers, e.g.

- https://www.tf-pm.org/resources/manifesto
- https://www.forbes.com.au/brand-voice/uncategorized/processminer-is-revolutionizing-the-manufacturing-industry-with-a-turn-key-ai-driven-predictive-and-prescriptive-analytic-platform/

Infomedia has deeper descriptions of the product and research papers in the automotive domain

Infomedia can provide initial event logs to start development.

Infomedia's Kim Horn has extensive experience in process mining, using Python packages, R packages like BupaR, and Research Tools like: ProM.

Kim Horn also has IP developed to extend RDR with Process models for exception-based extension, making improvement significantly easier.

For the team to understand the domain free versions of process mining tools includes ProM, Apromore, and PM4Py. ARIS Process Mining has a light version, ARIS Process Mining Element.

Note for Item 11 - below

3-4 groups are suggested – to work on different aspects together. One set could work on the e2e Protype function using Open-Source components. Another set could work on producing the code for the core Inductive Mining Algorithms only; to remove any issue we may have with Open-Source components. These could be swapped in/out of the Prototype.