

Supervisors:

Prof. Dr. Dimka Karastoyanova

Asst.Prof. Estefanía Talavera

Martínez, PHD

Presented By:

Pooja Gowda

(S4410963)

18<sup>th</sup> July 2022

#### **Contents**

- 1. Introduction
- 2. Research Question
- 3. All about the Data Set
- 4. Types of ProM Miners
- 5. Implementation
- 6. Results
- 7. Challenges
- 8. Discussion and Conclusion
- 9. Questions

#### Introduction:

- The involvement of data science and data mining in building process models has arisen.
- Advantages of data mining techniques is vast when combined with process mining.
- Objective?

#### **Research Questions**

Research questions are a sets of questions to summarize the objective of this project and the answers to these questions helps in formulating the project conclusion. The main questions were:

- 1. What are the existent Process Mining Techniques that can be applied to egocentric datasets for behavior discovery?
- 2. Can we apply Process Mining technique on extracted context information (location and objects) from egocentric videos for daily human behavior analysis?

#### **Data Set**

- Data set is downloaded from ego4d-data.org
- The data set is diverse in its geographic range, scenarios, participants, and captured modalities.
- Using various off-the-shelf head-mounted cameras: GoPro, Vuzix Blade, Pupil Labs,
  ZShades, OR- DRO EP6, iVue Rincon 1080, and Weeview.
- The egocentric data set is a data set captured from the first-person or "egocentric" point of view where we capture the activities with the respect to the eyes of a user directly.
- Types of ProM Miners can be applied on this dataset.

# Types of ProM Miners Algorithms

Different type of ProM miner algorithms are:

- 1. Alpha Miner
- 2. Inductive Petri Net
- 3. Evolutionary Tree Miner (ETMd)
- 4. Discovery Graph (Casual Net Mining)

#### 1. Alpha Miner

Alpha Miner is the algorithm that helps in filling the link between event logs and the process model.

 This algorithm helps in building process models based on event logs with relations and causalities between the steps of processes.

#### The Alpha miner generates:

- 1. A Petri net model with all the transitions being visible, unique, and correspond to the classified events.
- 2. The initial marking Describes the status of the Petri net model when the execution starts.
- 3. The final marking Describes the status of the Petri net model when the execution ends.

#### 2. Inductive Miner

The Splits between the start and final state of the process in a log are detected by using Inductive Miner algorithm.

Types of splits identified are as follows:

- I. Sequential
- II. Parallel
- III. Concurrent
- IV. Loop
  - On identifying the splits, this repeats until a base case is found.

# 3. Evolutionary Tree Miner (ETMd)

- It is a genetic algorithm that conducts the discovery process depending on the four specific quality dimensions from the discovered model which are:
  - I. Fitness
  - II. Precision
- III. Generalization
- IV. Complexity

#### 4. Discovery Graph

The causal relations extracted from the event log and the background knowledge on constraints over the topology of the produced models that are used.

# **Implementation**

Implementation was carried out in 3 stages:

- 1. Preprocessing
- 2. Converting CSV to XES
- 3. Apply ProM Miners / Algorithms.

# 1. Preprocessing

- Most of the algorithms require data with a limited amount of changes. Therefore, the data that is used remained unchanged as much as possible.
- The data was split into three separate CSV files concerning the video source with the specific case number to be 1,2, or 3(representing process occurring in different location).
- Further, these CSV files were merged into one to execute miners on this data.

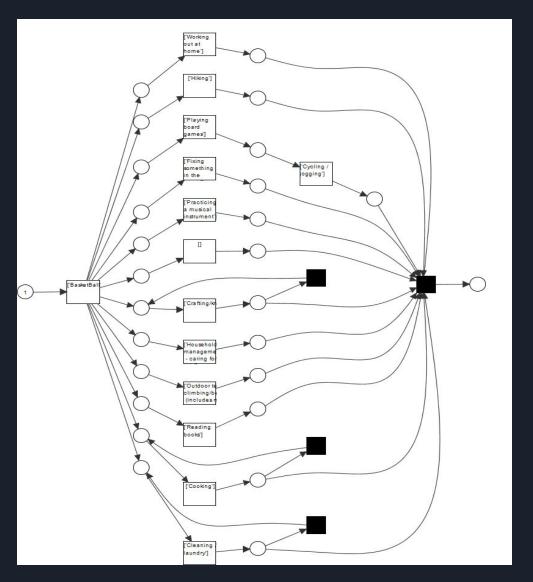
#### 2. Convert CSV to XES

- The parsed data is stored in CSV with case identifiers
- As per XES format having an Id and the case number is important for running the data in Prom.
- Prom provides a plugin that converts CSV to XES format, using this I get 'XESEventlog'.

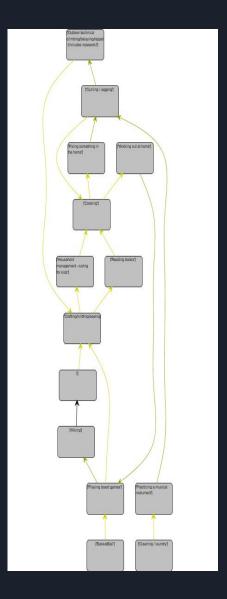
# 3. Apply Process Mining Techniques

- I opted to work on smaller subsets of data, due to the difficulty in running the mining algorithms several times.
- Further applied ProM miners and noted the results.

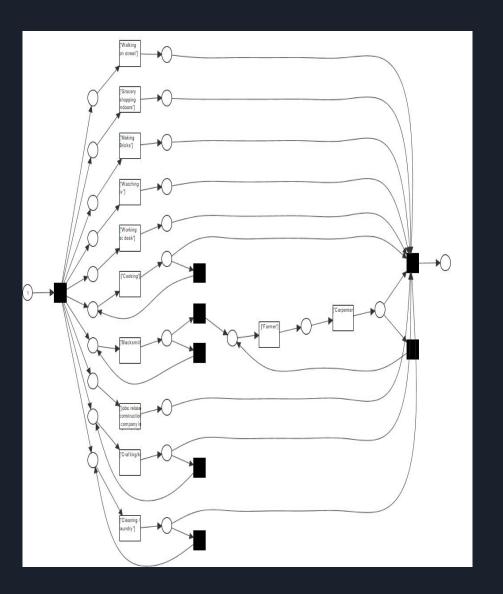
The resultant process model using Inductive miner on data logged from Bristol location(case3)



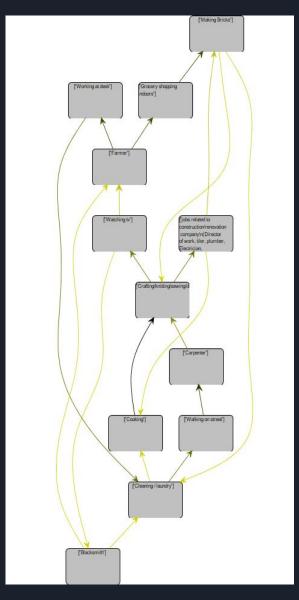
The resultant process model using discovery graph on data logged from Bristol location case(3)



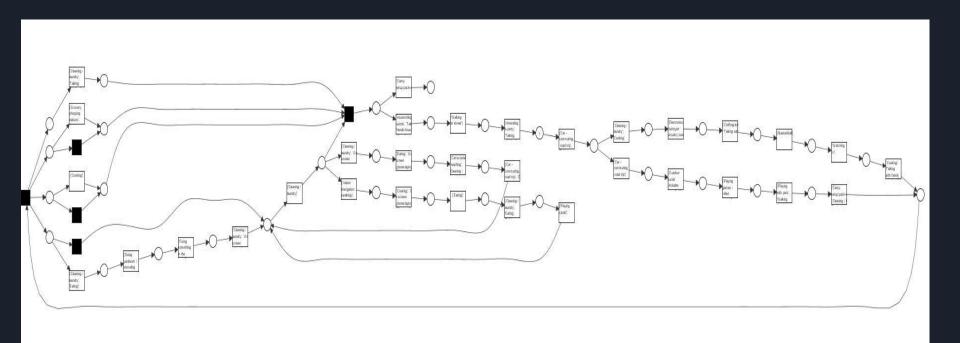
The resultant process model using Inductive miner on data logged from iiith location(case2)



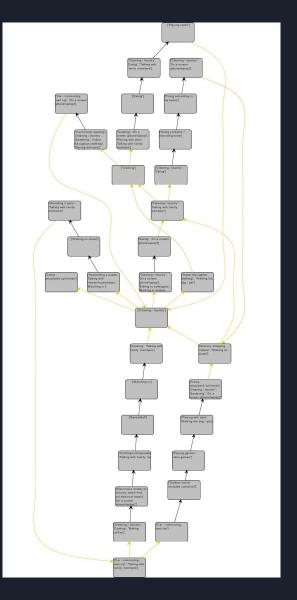
The resultant process model using discovery graph on data logged from iiith location case(2)



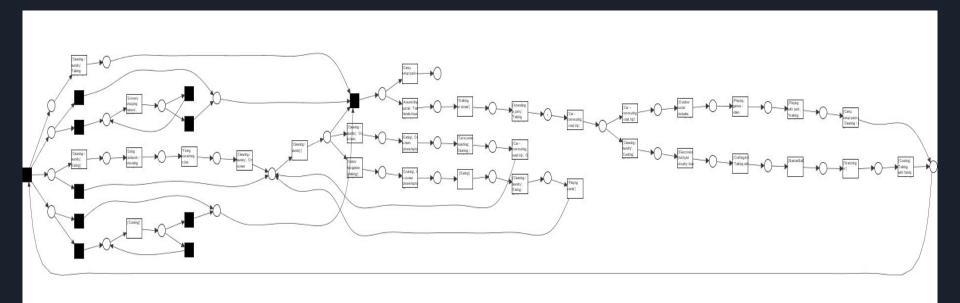
The resultant process model using inductive miner on data logged from minnesota location case(1) with noise 0.20



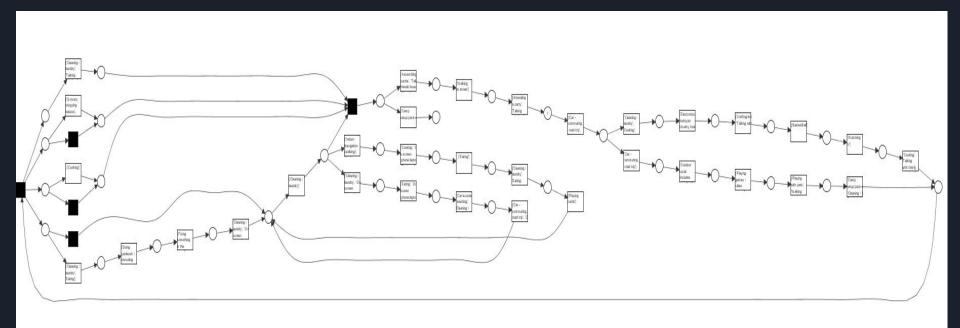
The resultant process model using discovery graph on data logged from minnesota location case(1) 17



The resultant process model using inductive miner on data logged from minnesota location case(1) with noise 0.10



The resultant process model using inductive miner on data logged from minnesota location case(1) with noise 0.30



# Challenges

- 1. **Data sets:** with time and date stamp (dd-mm-yyyy) would help to understand when the activity started and ended.
- 2. Using a several algorithm: Applying the several algorithms on the data set would definitely turn out to be a value added to the results.

#### Conclusion

- **1. Goal:** to research the different available methods by applying process mining where the data is egocentric.
- 2. Challenges: The obtained dataset had no timestamps on them, hence more time was spent in exploring the data.
- 3. **Result/Outcome:** A positive sign towards the potential of including process mining techniques over generally logged data.
- 4. Conclusion: There is potentiality in using ego-centric data sets to generate process models. Also, by going further in this research direction, we can discover the number of occurrences of a certain event which talks about the health status of the user and set of people living in that area.

#### References

[1] Wil van der Aalst.(2011). "Process Mining" Springer-Verlag Berlin Heidelberg, Second Edition. ISBN 978-3-662-49851-4

[2] Cristina-Claudia DOLEAN.(2014) "Mining Product Data Models: A Case Study" Informatica Economică vol. 18

# **Questions?**