

Akademia Górniczo-Hutnicza w Krakowie

*Wydział Inżynierii Metali i Informatyki Przemysłowej*

Kierunek: Informatyka Techniczna

**Praca dyplomowa**

**Magisterska**

**Waldemar Świder**

*Analiza szeregów czasowych w zastosowaniu do rozpoznawania złożonych aktywności w danych zawierających akcje podejmowane przez użytkownika systemu smart home*

*Time series analysis used to recognize complex activities in data containing actions of a smart home user*

**Promotor: dr inż. Gabriel Rojek**

Spis treści

[1. Wstęp 3](#_Toc161554958)

[2. Cel pracy 5](#_Toc161554959)

[3. Przegląd literatury i istniejących rozwiązań 6](#_Toc161554960)

[Taksonomia metod HAR (eng. Human Activity Recognition) 6](#_Toc161554961)

[Metody uczenia głębokiego 7](#_Toc161554962)

[Modelowanie tematyczne 7](#_Toc161554963)

[CEP (eng. Complex Event Processing) 8](#_Toc161554964)

[4. Testy 10](#_Toc161554965)

[5. Wnioski 11](#_Toc161554966)

[6. Podsumowanie 12](#_Toc161554967)

[Bibliografia 13](#_Toc161554968)

[Wykaz tabel i wykresów 14](#_Toc161554969)

[Załączniki 15](#_Toc161554970)

# Wstęp

W dobie dominacji sztucznej inteligencji jako technologii, w której upatruje się rozwiązania wszelkich problemów ludzkich, pojawia się oczekiwanie implementacji inteligentnych rozwiązań w każdej ze sfer życia. W przypadku Internetu rzeczy *(eng. Internet of Things / IoT)*, czy tak zwanych inteligentnych domów *(eng. Smart Home)* trwają już pewne prace nad wdrożeniem takich rozwiązań, aczkolwiek nadal nie do końca spełniają one założenia systemów inteligentnych, ponieważ w większości wypadków, działają one jedynie w odpowiedzi na bezpośrednie polecenia, bądź predefiniowane wyzwalacze i nie podejmują decyzji w sposób samodzielny.



Rysunek a

W przypadku istniejących na rynku systemów typu „Smart Home”, często mamy do czynienia z ubogą ofertą aspektu „Smart” w ramach platform. Przyczyn tego może być kilka, jak na przykład brak ujednoliconego standardu komunikacji między urządzeniami różnych producentów, brak jednoznacznego pomysłu na działanie tego typu rozwiązań, czy też zwyczajnie kwestia kosztów dodatkowych udogodnień.

Inteligentne systemy[1] mogą być opisane jako rozwiązanie, będące w stanie, w sposób nienadzorowany, uczyć się na podstawie archiwalnych danych, wyciągać wnioski i samodzielnie podejmować decyzje. Taka technologia jest wykorzystywana do automatyzacji powtarzalnych zadań. Problematyka, która zostanie w tej pracy poruszona, dotyczy automatycznego wykrywania zadań, które to można powierzyć automatom. Zadania takie przedstawione byłyby w formie makr – listy prostych instrukcji, składających się na pewną złożoną czynność.

W przypadku branży inteligentnych domów, możliwym byłoby rozwinięcie istniejących technologii, o modele przewidywania aktywności użytkownika, celem wyręczenia go z rutynowych czynności. Trudnym problemem wydaje się w tym wypadku precyzyjna predykcja, kiedy dokładnie użytkownik wykona konkretną czynność (na przykład nastawi wodę na poranną kawę), ponieważ użytkownik może nie posiadać stałych pór dla zadanych czynności, a także jego plan dnia może podlegać stałym modyfikacjom. W tym wypadku właściwym podejściem wydaje się być asocjacyjna analiza tych „atomowych” czynności. Ideą tego podejścia, jest obserwacja mówiąca, że każda złożona czynność daje się określić przez zestaw prostszych instrukcji, aż do tych najprostszych „atomowych”, które to łatwo precyzyjnie zdefiniować i zautomatyzować. Można więc na podstawie kontekstu wykonywanych czynności prostych i innych wykrytych danych (np. pomiarów z czujników), przewidywać zamiary użytkownika i ubiegać go w nich.

Aby rozwiązanie opisane powyżej było kompletne, system taki musi w sposób nie wymagający ingerencji użytkownika, wykrywać schematy w jego przeszłych działaniach i samodzielnie układać makra, które to następnie powinien aplikować we właściwych sytuacjach.

Warto zauważyć, że rozwiązanie dla tego problemu w branży Smart Home, byłoby możliwe do zaaplikowania również w innych dziedzinach, ponieważ problematyka przewidywania akcji użytkownika, oraz makra jako narzędzie optymalizacji procesów, występują praktycznie zawsze gdy mamy do czynienia z aktywnością ludzką. Przykładem powyższego, może być dziedzina UX, w której za cel stawia się jak największe uproszczenie i uwygodnienie interakcji użytkownika z systemem, tutaj istotnym jest przewidzieć oczekiwania i kolejne akcje użytkownika, aby ubiec go i spełnić jego potrzeby zanim sam będzie w stanie je określić. Innym przejawem takich predykcji, są systemy sprzedażowe biletów lotniczych dostępne przez Internet, w ich wypadku dochodzi do predykcji oczekiwań użytkownika wobec ceny i ewentualnych destynacji, które następnie są przekazywane w ramach systemu podpowiedzi.

Wobec powyższych, można śmiało stwierdzić, że na rynku istnieje zapotrzebowanie na rozwiązanie problemu postawionego w temacie pracy, a wyniki badań mogą być użyteczne w szerokim spektrum zastosowań.

# Cel pracy



Rysunek b

Celem niniejszej pracy jest dokonanie dogłębnej analizy szeregów czasowych w zastosowaniu do rozpoznawania złożonych aktywności użytkownika w obrębie dziedziny Internetu rzeczy i systemów typu *smart home,* oraz wypracowanie metodologii użytecznej w rzeczywistych aplikacjach. Poniżej przedstawione zostały najważniejsze punkty, które zostaną poruszone w ramach realizacji tematu pracy:

* **Opracowanie danych.**

W tej części omówione zostaną struktury danych wykorzystywanych do analizy szeregów czasowych i rozpoznawania złożonych aktywności. Poruszone zostaną też metody ich oczyszczania i preprocesowania dla modeli i algorytmów wykorzystywanych w dalszych częściach pracy.

* **Analiza szeregów czasowych.**

Przedłużeniem opracowania danych oraz wstępnym krokiem do detekcji złożonych zachowań, jest analiza zbioru danych. Na tym etapie omówione zostaną algorytmy uczenia maszynowego oraz techniki detekcji trendów i istotnych cech w zadanych zbiorach danych. Głównym celem tej części jest wypracowanie ogólnej charakterystyki zachowań użytkowników.

* **Rozpoznawanie złożonych aktywności.**

Część poświęcona zostanie właściwemu procesowi wykrywania złożonych aktywności na podstawie wniosków podjętych w poprzedzających rozdziałach. Jej celem będzie dobór i opracowanie algorytmów oraz optymalizacji.

* **Wdrożenie, walidacja i porównanie modeli.**

W ramach podsumowania wniosków podjętych w dotychczasowych rozdziałach, dokonana zostanie implementacja i porównanie wybranych modeli. Uzyskane wyniki posłużą do wyboru tych o najlepszym dopasowaniu do celu pracy.

Jak opisano powyżej, praca będzie mieć charakter opisu procesu, który został podzielony na wymienione etapy; będą się one opierać o stopniowe podnoszenie poziomu abstrakcji danych wejściowych.

# Przegląd literatury i istniejących rozwiązań

Źródła dla tej pracy zostały dobrane poprzez przeszukanie dostępnych baz prac naukowych z użyciem słów kluczowych, które zostaną w poniższej pracy poruszone, takich jak „analiza szeregów czasowych”, czy też „rozpoznawanie zachowań w systemach IoT”. Przegląd polegać będzie na zwięzłym przedstawieniu ich treści i analiza w odniesieniu do tematyki tejże pracy. Celem tego rozdziału jest omówienie istniejących już rozwiązań, dotarcie do aktualnego stanu wiedzy dla badanej branży i wypracowanie własnych wniosków na temat tego w jakim kierunku powinny zmierzać prace nad rozwiązaniem problemu postawionego w celu tejże pracy.

## Taksonomia metod HAR *(eng. Human Activity Recognition)*

W przypadku rozpoznawania aktywności ludzkiej istnieją dwa główne rodzaje metod: jednomodalne i wielomodalne[2]. Jednomodalne metody opierają się o pracę na danych z pojedynczego źródła, lub/i o zbliżonej strukturze, natomiast w przypadku wielomodalnych, analiza dokonywana jest na różnorodnych danych pochodzących z wielu źródeł. Ów podział rozbija się dalej na podrodzaje metod, z czego w przypadku IoT najwłaściwszymi zdają się być poniższe metody jednomodalne:

* **Metody przestrzenno-czasowe.**

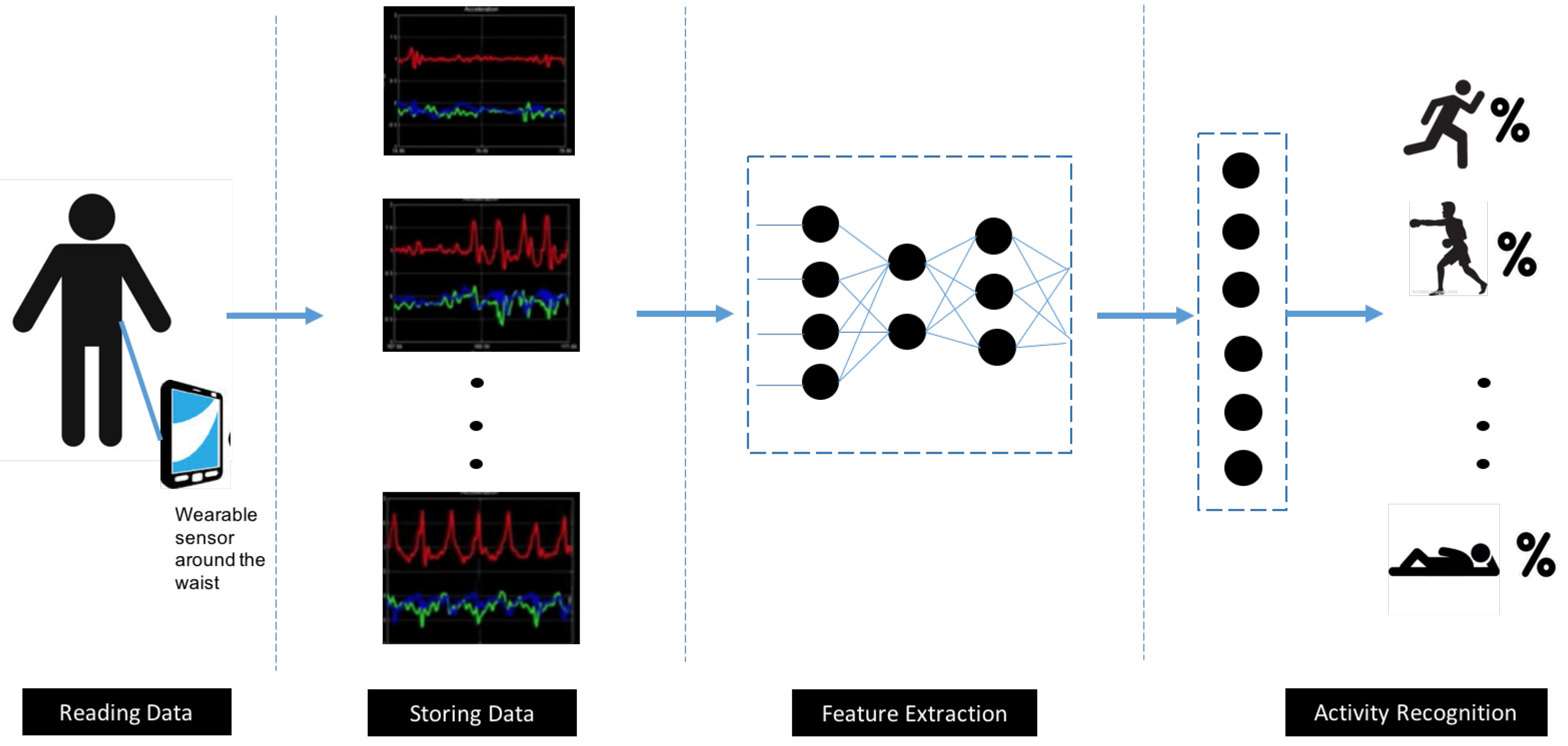
Pozwalają na rozpoznanie aktywności na podstawie cech przestrzenno-czasowych. Stosuje się je przede wszystkim przy analizie danych typu wideo i pochodzących z akcelerometrów. Ich wadą jest niska skuteczność w przypadku aktywności o niskiej dynamice.

* **Metody stochastyczne.**

Metody te wykorzystują modele statystyczne takie jak warunkowe pola losowe (CRF), czy też ukryte modele Markowa (HMM) do rozpoznawania zachowań ludzkich. Są one dość skuteczne w modelowaniu danych pochodzących z szeregów czasowych, ale wymagają dużych zbiorów treningowych.

* **Metody oparte na regułach.**

Wykorzystują predefiniowanych zbiorów reguł i atrybutów do rozpoznawania aktywności. Są one wyjątkowo skuteczne w wykrywaniu konkretnych i łatwo-opisywalnych czynności, ale brak im elastyczności i nie obsługują one niezdefiniowanych uprzednio aktywności.

Biorąc pod uwagę potrzebę analizy z różnych rodzajów czujników urządzeń, niekoniecznie o charakterze przestrzenno-czasowym, a także zamiar rozpoznawania aktywności w sposób nienadzorowany, drogą eliminacji, najwłaściwszymi z metod dla problemu postawionego w tej pracy, pozostają metody stochastyczne, ze względu na ich elastyczność i wysoką skuteczność w rozpoznawaniu zachowań ludzkich w nieregularnych zbiorach danych.  


Rysunek c

## Metody uczenia głębokiego

Biorąc pod uwagę dużą popularność uczenia głębokiego, należy przeanalizować również przydatność tychże metod w przypadku detekcji aktywności w danych pochodzących z systemów IoT. Spośród wielu, przeanalizowane[3] zostało kilka wybranych, najlepiej przystosowanych metod:

* **Rekurencyjne Sieci Neuronowe (RNN).**

Osiągają wysokie wyniki w przypadku analizy szeregów czasowych i nadają się do tworzenia prognoz. Ich wadą jest natomiast słaba detekcja zależności długoterminowych.

* **Sieci Long Short-Term Memory (LSTM).**

Stanowią ulepszoną, pod kątem detekcji zależności długoterminowych, wersję RNN. Odpowiednie do detekcji trendów i cech w danych o dużej rozpiętości czasowej.

* **Sieci neuronowe oparte na transformerach.**

Sprawdzają się w przypadku zadań typu sequence-to-sequence, oraz w wizyjnych systemach komputerowych. Ich oderwana od sekwencyjnego przetwarzania (typowego dla metod RNN i LSTM) architektura, pozwala na zastosowanie przy obliczeniach równoległych.

* **Maszyny Boltzmanna (BM) i Głębokie Sieci Wierzeń (DBN).**

Odpowiednie do zastosowań w uczeniu nienadzorowanym, jak na przykład ekstrakcja cech i klasyfikacja bez określonych etykiet. Sprawdzają się w eksploracji charakterystyk danych IoT, pozwalając na wykrywanie zależności w strukturach tychże danych.

Biorąc pod uwagę charakterystyki wyżej wymienionych metod, w przypadku zastosowania do ekstrakcji cech z szeregów czasowych o dużej rozpiętości czasowej, najodpowiedniejszymi są metody LSTM, BM i DBN. Metody te jednocześnie są metodami stochastycznymi co zbiega się z wnioskami podjętymi w poprzednim punkcie i nadaje pewien kierunek w doborze metod.

## Modelowanie tematyczne

**Modelowanie tematyczne** opisuje zbiory danych jako dokumenty, z których każdy może zawierać jeden lub więcej "tematów" reprezentujących, na przykład w przypadku danych IoT, różne rodzaje aktywności. Modele, jak na przykład Latent Dirichlet Allocation (LDA), jest następnie wykorzystywany do rozpoznawania tych tematów w sposób nienadzorowany, co oznacza, że model sam identyfikuje potencjalne tematy (cechy) bez konieczności manualnego oznaczania danych przez ludzi.

W badaniach[6][7], modelowanie tematyczne zostało wykorzystane do analizy danych z czujników, celem identyfikacji wzorców odpowiadających różnym codziennym rutynom, takim jak podróż do pracy, praca, posiłki czy też odpoczynek. Poprzez analizę aktywności w różnych przedziałach czasowych, model jest w stanie nauczyć się i zidentyfikować powtarzające się schematy, które mogą zostać sklasyfikowane jako specyficzne rutyny lub zwyczaje.

Modelowanie tematyczne opiera się na statystycznej analizie rozkładu słów lub, w jak w zadanym przypadku systemów IoT, rozkładu występowania różnych aktywności w dokumentach reprezentujących dane pochodzące z szeregów czasowych. Kluczowe w tym procesie jest zastosowanie algorytmów takich jak LDA do estymacji, tematów o najwyższym prawdopodobieństwie wystąpienia dla danego zestawu danych oraz detekcji aktywności najbardziej charakterystycznych dla każdego z tematów.

W pracy "Discovery of Activity Patterns using Topic Models"[6], modelowanie tematyczne zostało wykorzystane do rozpoznawania codziennych rutyn jako probabilistycznej kombinacji wzorców aktywności. Pozwoliło to w sposób nienadzorowany, na automatyczne odkrywanie tychże wzorców w codziennej rutynie użytkownika​​. Podobne podejście zostało zaprezentowane w dokumencie „Discovering Routines from Large-Scale Human Locations using Probabilistic Topic Models”, gdzie podejście do rozpoznawania cech w ramach dużych zbiorów danych również opierało się na podobnych zasadach analizy wzorców aktywności i ich automatycznej klasyfikacji.

## CEP *(eng. Complex Event Processing)*

Complex Event Processing (CEP) to technologia przetwarzania danych w czasie rzeczywistym, która umożliwia monitorowanie, analizowanie i reagowanie na zdarzenia pochodzące z różnych źródeł. W kontekście Internetu Rzeczy (IoT), CEP odgrywa kluczową rolę w wykrywaniu wzorców i anomalii w zbiorach danych o dużej objętości.

CEP przetwarza dane pochodzące z różnych źródeł, takich jak czujniki temperatury, wilgotności, ruchu czy zużycia energii. Dane są pozyskiwane z wysoką częstotliwością, co wymaga odpowiedniego ich przygotowania. Przygotowanie to obejmuje odfiltrowanie zakłóceń i nieistotnych informacji, agregację zdarzeń pokrewnych, oraz wykrywanie wzorców zachowań.

W badanych rozwiązaniach[8] wykorzystana została rozproszona architektura CEP, która pozwala na przetwarzanie danych blisko ich źródła (tzw. edge computing), redukując obciążenie sieci i serwerów backendowych. W architekturze tej wyróżnia się dwa główne komponenty: preprocessingu (po stronie klienta) i wnioskowania (po stronie serwera). Pozwala to na efektywne przetwarzanie dużych zbiorów danych, przy jednoczesnej minimalizacji obciążenia sieci, oraz serwerów.

CEP jest przydatny w systemach IoT ze względu na swoją zdolność do szybkiego przetwarzania znacznych zbiorów danych uzyskiwanych przez duże systemy czujników w czasie rzeczywistym. Może być on więc z powodzeniem wykorzystywany w systemach typu Smart Home, gdzie wydajność obliczeniowa i sieciowa są szczególnie istotne.

## Las Losowy

Algorytm Lasu Losowego[9] *(eng. Random Forest)* to technika uczenia maszynowego, wykorzystująca agregację wielu drzew decyzyjnych do generowania bardziej stabilnych i dokładnych predykcji. Celem zastosowania Lasu Losowego w kontekście wykrywania cech jest zwiększenie zdolności modelu do identyfikacji i różnicowania istotnych atrybutów w zestawie danych, co może być kluczowe w wielu aplikacjach, takich jak rozpoznawanie wzorców, klasyfikacja obiektów czy analiza predykcyjna. Wykorzystanie Random Forest w wykrywaniu cech pozwala na poprawę dokładności i efektywności modeli predykcyjnych przez skuteczną selekcję i ocenę znaczenia atrybutów.

# Opracowanie danych

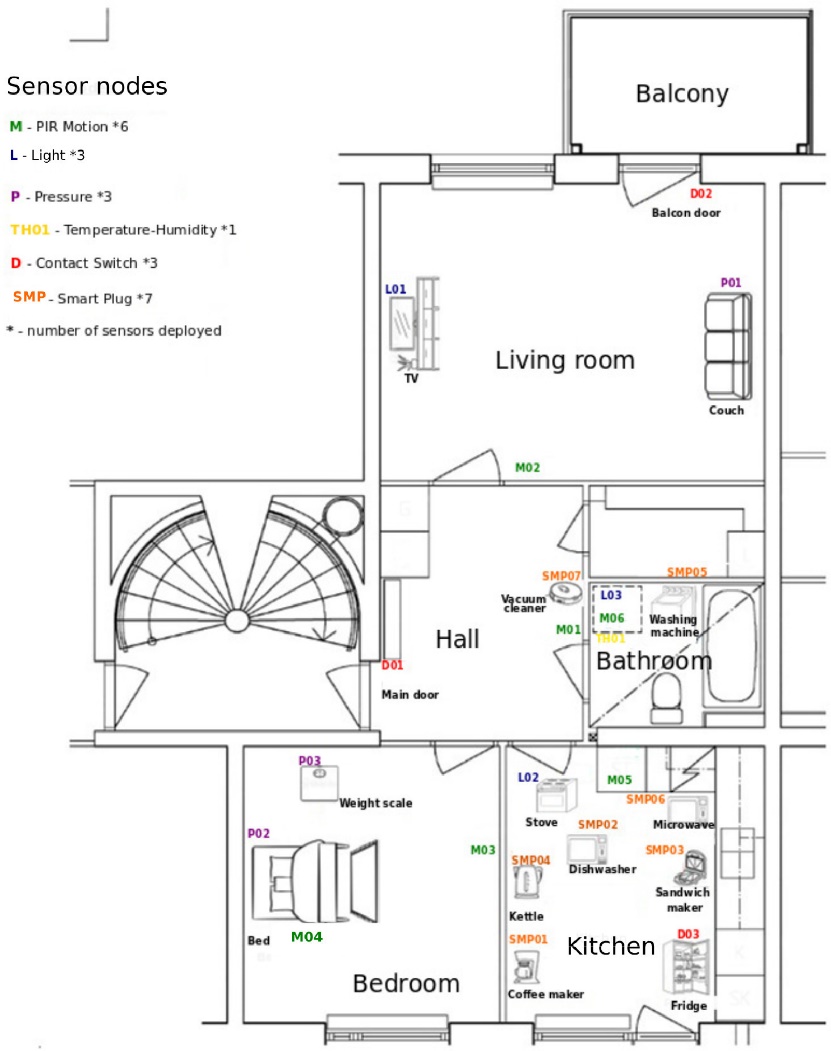
## Opis zbioru danych

Zbiorem danych, który zostanie wykorzystany do badań, będzie zestaw szeregów czasowych danych pochodzących z sensorów inteligentnego domu[d]. Zawiera on dane na temat aktywności różnych urządzeń, a także informacje o wartościach parametrów środowiskowych jak temperatura, czy wilgotność powietrza.

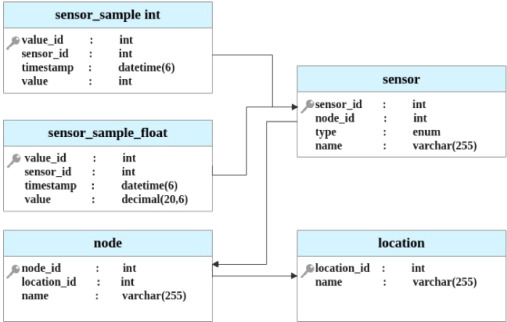
Zgromadzone rekordy zostały uzyskane na przestrzeni 6 miesięcy ciągłych pomiarów z częstotliwością 1Hz. Pochodzą z urządzeń i czujników zaprezentowanych w tabeli/wykresie poniżej:

| **Sensor** | **Sensor Type** | **Object Monitored** | **Location** |
| --- | --- | --- | --- |
| M01 | Motion | Ambience | Corridor |
| M02 | Motion | Ambience | Living room |
| M03 | Motion | Ambience | Bedroom |
| M04 | Motion | Ambience | Bedroom |
| M05 | Motion | Ambience | Kitchen |
| M06 | Motion | Ambience | Bathroom |
| L01 | Light | TV | Living room |
| L02 | Light | Stove | Kitchen |
| L03 | Light | Ambience | Bathroom |
| P01 | Pressure | Couch | Living room |
| P02 | Pressure | Bed | Bedroom |
| P03 | Pressure | Weight scale | Bedroom |
| TH01 | Temperature & Humidity | Ambience | Bathroom |
| D01 | Reed switch | Door contact | Entrance |
| D02 | Reed switch | Door contact | Balcon |
| D03 | Reed switch | Fridge Door contact | Kitchen |
| SMP01 | Smart plug | Coffee maker | Kitchen |
| SMP02 | Smart plug | Dishwasher | Kitchen |
| SMP03 | Smart plug | Sandwich maker | Kitchen |
| SMP04 | Smart plug | Kettle | Kitchen |
| SMP05 | Smart plug | Washing machine | Bathroom |
| SMP06 | Smart plug | Microwave | Kitchen |
| SMP07 | Smart plug | Vacuum cleaner | Corridor |

Tabela I przedstawiająca wykorzystane w poborze danych urządzenia i czujniki

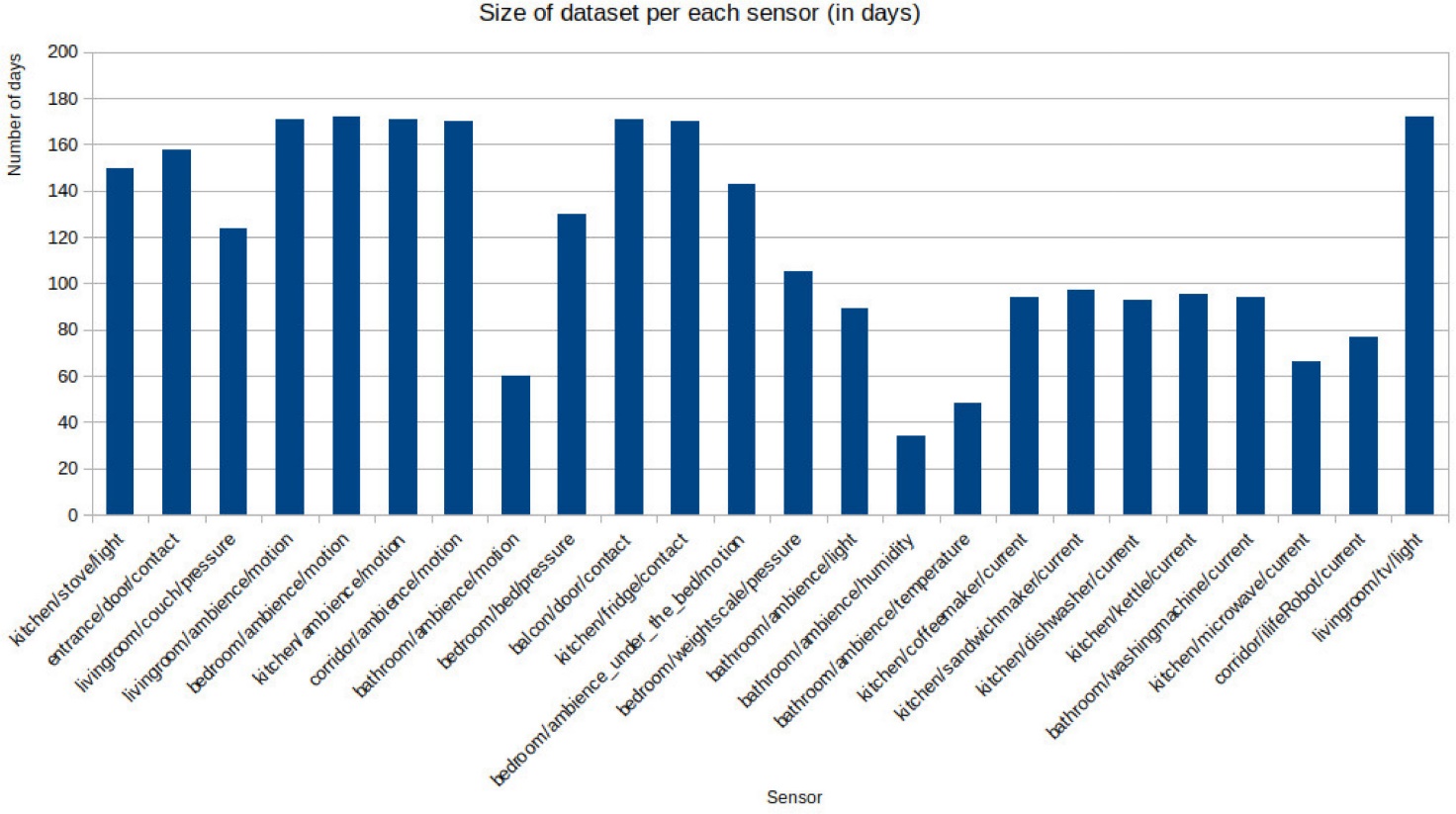


Rysunek e – plan domu w którym dokonano pomiarów



Wykres II – plan domu w którym dokonano pomiarów

Czujniki były wprowadzane do systemu w stopniowo, co skutkuje jedynie częściowym pokryciem okresu pomiarów, przez część urządzeń. Poniżej zaprezentowany został wykres przedstawiający to obłożenie czasowe dla każdego z czujników:



Wykres III

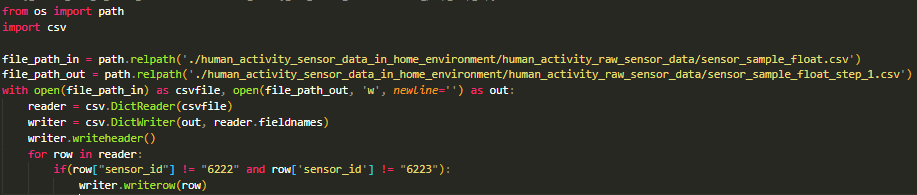
## Oczyszczenie danych

Pomimo niepełnego pokrycia półrocznego okresu przez większą część czujników, zbiór pozostaje znacznych rozmiarów (około 12.4GB rekordów danych), co utrudnia jego analizę i przetwarzanie. Rozmiar ten wynika przede wszystkim z wysokiej częstotliwości pozyskiwania danych i co za tym idzie, dużego odsetku rekordów nie zmieniających wartości względem poprzedniego odczytu.

Do celów tej pracy potrzebne będą jedynie dane dotyczące atomowych czynności, a więc informacje o rozpoczęciu, bądź zakończeniu prostej czynności, którą możemy wywnioskować na podstawie pojedynczego odczytu.

### Eliminacja zbędnych czujników/urządzeń

Aby sprowadzić dane do tej formuły, należy odrzucić dane o temperaturze i wilgotności powietrza jako tych, które nie informują w jednoznaczny sposób o wykonywanej czynności.

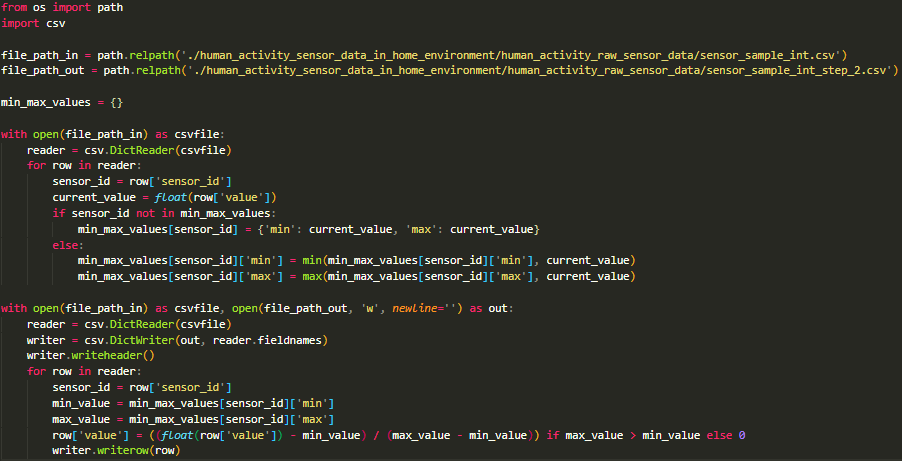


Załącznik f – kod pierwszego etapu oczyszczenia danych

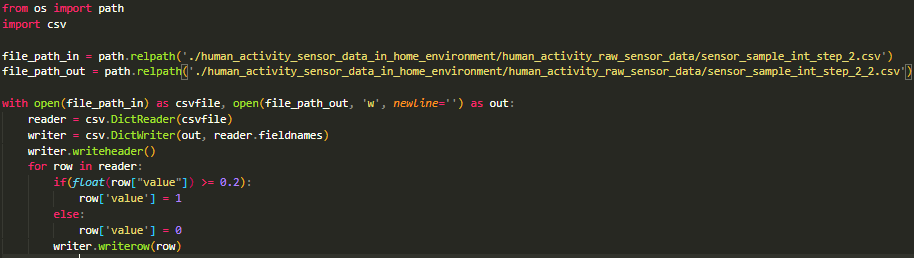
Wielkość danych po pierwszym etapie oczyszczenia: 12.4GB -> 11.66GB (-0.74GB).

### Binaryzacja danych

Kolejnym krokiem będzie binaryzacja danych. Ponieważ istotnym dla celów badawczych jest jedynie rozpoznanie stanów użytkowania i nieaktywności urządzeń, dane zostaną sprowadzone do wartości binarnych, gdzie wartość zera informuje o nieaktywności danego urządzenia/braku detekcji aktywności, zaś jeden wskazuje na aktywność wokół danego urządzenia/czujnika. W tym celu w pierwszej kolejności dokonana zostanie normalizacja odczytów do zakresu 0-1, aby dalej na podstawie progu wynoszącego 20% wartości maksymalnej (określonego na podstawie obserwacji charakterystyki danych), sklasyfikować binarnie dane.



Załącznik g – kod drugiego etapu oczyszczenia danych (normalizacja)

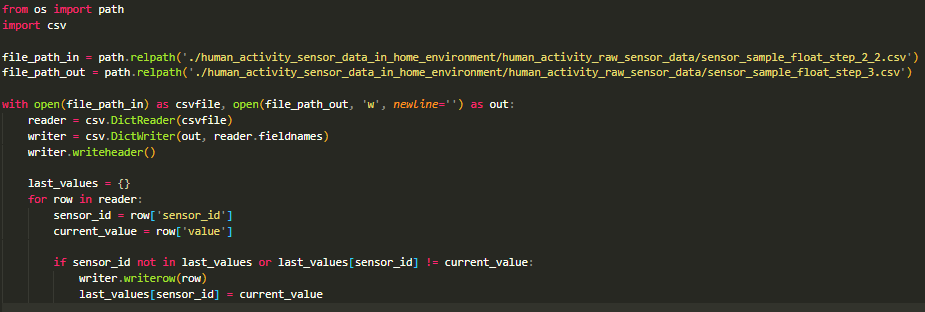


Załącznik h – kod drugiego etapu oczyszczenia danych (binaryzacja)

Wielkość danych po drugim etapie oczyszczenia: 11.66GB -> 9.96GB (-1.7GB).

### Eliminacja powtarzających się rekordów

Następnie usunięte zostaną powtarzające się rekordy o tych samych wartościach, w ramach tych samych czujników. W ten sposób w zbiorze danych pozostaną jedynie rekordy informujące o rozpoczęciu i zakończeniu czynności.

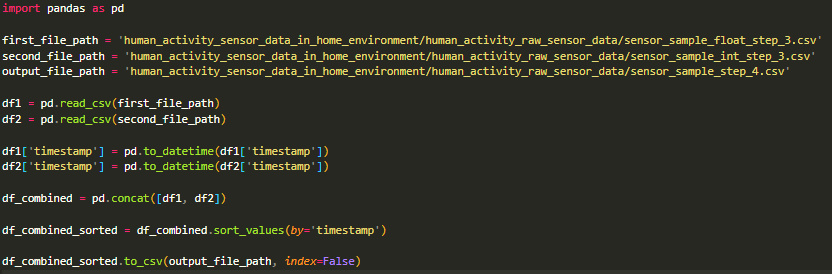


Załącznik i – kod trzeciego etapu oczyszczenia danych (eliminacja kolejnych duplikacji)

Wielkość danych po trzecim etapie oczyszczenia: 9.96GB -> 0.09GB (-9.87GB).

### Połączenie plików

Ostatnim etapem przygotowania danych będzie połączenie dwóch plików dotyczących danych pierwotnie całkowitych i ułamkowych.



Załącznik j – kod połączenia obu plików z pomiarami

Ostatecznie uzyskany zbiór danych pomiarowych ma wielkość 88.6MB, a więc poniżej 1% wielkości danych wejściowych. Dane te, zbinaryzowane i oczyszczone ze zbędnych duplikacji posłużą do detekcji charakterystyk zachowań z użyciem odpowiednich modeli.

# Testy

# Wnioski

# Podsumowanie

# Bibliografia

[1] Imre J. Rudas, János Fodor; “Intelligent Systems”; Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844 Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 132-138

[2] Michalis Vrigkas, Christophoros Nikou, Ioannis A. Kakadiaris; “A Review of Human Activity Recognition Methods”; Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece, Computational Biomedicine Laboratory, Department of Computer Science, University of Houston, Houston, TX, USA; 16.11.2015

[3] Kuruva Lakshmanna, Rajesh Kaluri, Nagaraja Gundluru, Zamil S. Alzamil, Dharmendra Singh Rajput, Arfat Ahmad Khan, Mohd Anul Haq, and Ahmed Alhussen; „A Review on Deep Learning Techniques for IoT Data”; Vellore Institute of Technology (VIT), Vellore 632014, India; lakshman.kuruva@vit.ac.in (K.L.); rajesh.kaluri@vit.ac.in (R.K.); nagaraja.g@vit.ac.in (N.G.) 2 Department of Computer Science, College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia; z.alzamil@mu.edu.sa 3 School of Manufacturing Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand 4 Department of Computer Engineering, College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia \* Correspondence: dharmendrasingh@vit.ac.in (D.S.R.); khansatwat@gmail.com (A.A.K.); m.anul@mu.edu.sa (M.A.H.); aa.alhussen@mu.edu.sa (A.A.

[4] Liming Chen, Chris D. Nugent, and Hui Wang ; „A Knowledge-Driven Approach to Activity Recognition in Smart Homes”; IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 6, JUNE 2012

[5] Zachary Menter , Wei Zhong Tee , Rushit Dave; „Application of Machine Learning Based Pattern Recognition in IoT Devices: Review”; University of Wisconsin-Eau Claire, Eau Claire WI 54701, USA

[6] Tam Hu ˆ ynh, Mario Fritz and Bernt Schiele; „Discovery of Activity Patterns using Topic Models”; Computer Science Department TU Darmstadt, Germany

[7] KATAYOUN FARRAHI and DANIEL GATICA-PEREZ; „Discovering Routines from Large-Scale Human Locations using Probabilistic Topic Models”; IDIAP Research Institute Ecole Polytechnique Fed´ erale de Lausanne (EPFL)

[8] Ching Yu Chen, Jui Hsi Fu, Today Sung, Ping-Feng Wang, Emery Jou, Ming-Whei Feng; „Complex Event Processing for the Internet of Things and its Applications”; 2014 IEEE International Conference on Automation Science and Engineering (CASE) Taipei, Taiwan, August 18-22, 2014

[9] Benjamin Goehry, Hui Yan, Yannig Goude, Pascal Massart, Jean-Michel Poggi; „RANDOM FORESTS FOR TIME SERIES”; Laboratoire de Mathématiques d’Orsay, CNRS, Université Paris-Saclay, Faculté des Sciences d’Orsay, Bâtiment 307, 91405 Orsay, France, EDF Lab, 7 bd Gaspard Monge, 91120 Palaiseau, France, EDF Lab & Laboratoire de Mathématiques d’Orsay, CNRS, Université Paris-Saclay, Orsay, France, Laboratoire de Mathématiques d’Orsay, CNRS, Université Paris-Saclay, Orsay, France

# Wykaz tabel i wykresów

1. Tabela zbioru urządzeń wykorzystanych w gromadzeniu danych wykorzystanych w badaniach (załącznik d). <https://www.sciencedirect.com/science/article/pii/S2352340920315122>
2. Wykres struktury danych (załącznik d) <https://ars.els-cdn.com/content/image/1-s2.0-S2352340920315122-gr1.jpg>
3. Wykres czasowego obłożenia czujników (załącznik d) <https://ars.els-cdn.com/content/image/1-s2.0-S2352340920315122-gr2_lrg.jpg>

# Załączniki

1. Przykładowy schemat systemu Smart Home: https://ajeevi.com/wp-content/uploads/2023/06/1-15.jpg
2. Ilustracja analizy wykresów: <https://editor.analyticsvidhya.com/uploads/3951420200902_blog_-forecasting-with-time-series-models-using-python_pt2_website.png>
3. Przykładowy schemat rozwiązania HAR: <https://pub.mdpi-res.com/sensors/sensors-21-02141/article_deploy/html/images/sensors-21-02141-g001.png?1616391439>
4. Zbiór danych pochodzących z systemu Smart Home „Multi-sensor dataset of human activities in a smart home environment”: <https://www.sciencedirect.com/science/article/pii/S2352340920315122>
5. Plan rozkładu urządzeń czujników w inteligentnym domu, w którym zebrane zostały dane pomiarowe: <https://ars.els-cdn.com/content/image/1-s2.0-S2352340920315122-gr4.jpg>
6. Kod pierwszego etapu przygotowania danych „data\_prep\_step\_1\_clean.py”
7. Kod drugiego etapu przygotowania danych „data\_prep\_step\_2\_normalise.py”
8. Kod drugiego etapu przygotowania danych „data\_prep\_step\_2\_binarise.py”
9. Kod drugiego etapu przygotowania danych „data\_prep\_step\_3\_distinct.py”
10. Kod drugiego etapu przygotowania danych „data\_prep\_step\_4\_combine.py”