Universiteit Leiden
The Netherlands

Bachelor DSAI

Added value of advanced machine learning over baseline association rules mining

Anna Cieloszczyk

s3632571

Supervisors:
Prof.dr. Marco Spruit & Dr. Armel Lefebvre

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

12/08/2025

Abstract

The increasing prescription opioid usage requires new approaches that will be able to find meaningful, interpretable patterns in medical data. This study aims to apply Truly Unordered Rule Sets (TURS), an advanced machine learning model, and compare it with two baseline association rule mining approaches, Apriori and FP-Growth. The purpose of this study is to better understand the associations between social and psychological aspects in opioid prescriptions. To achieve that, Dutch GP medical records are analyzed using the ELAN dataset. TURS results are evaluated using metrics such as ROC AUC, coverage, average rule length, and generalization gap. Apriori and FP-Growth have been evaluated with support, confidence, lift, and conviction. The results show that TURS creates compact, non-overlapping rules that generalize well, but the probabilities for individual rules are low because of the multiclass nature of the data. While Apriori and FP-Growth retrieve many frequent patterns, MDL-based TURS offers a probabilistic approach that focuses on compressing the data. In alignment comparison, it uncovered 9 new associations within the data. Those rules show associations between the prescription of Tramadol and problems with partner's illness and transient stress response, as well as a corelation between the prescription of Morphine, Buprenorphine, and Codeine with sleep disorders. Both studies showed a correlation between memory disorders and the prescription of Tramadol. Although promising, TURS proves to not perform best with large datasets with unbalanced class distributions. Therefore, it may not yet be ready for application in large healthcare settings.

Contents

Abbreviations

ARM Association Rule Mining

ATC Anatomical Therapeutic Chemical classification system

CRISP-DM Cross Industry Standard Process for Data Mining

ELAN Extramural Leiden Academic Network Population Database

GP General Practitioner

ICPC International Classification of Primary Care

MDL Minimum Description Length

ML Machine Learning

OUD Opioid Use Disorder

TURS Truly Unordered Probabilistic Rule Sets

1 Introduction

1.1 Research motivation

The rapid increase in opioid use can be classified as one of the important public health challenges of the 21st century. The increase in opioid usage is particularly visible in the United States. However, such worrying tendencies are also present in Europe. In the Netherlands alone, the number of prescription opioid users nearly doubled from 4109 per 100,000 inhabitants to 7489 per 100,000 inhabitants between 2008 and 2017 (Kalkman, Kramers, van Dongen, van den Brink, & Schellekens, 2019). This rapid increase in just 9 years has been mainly caused by the increase in oxycodone prescriptions, which quadrupled from 574 to 2568 per 100000 inhabitants. Due to this, the number of opioid-related hospital admissions has tripled, and treatment for non-heroin opioid disorders nearly doubled between 2008 and 2015. Between 2008 and 2014 opioid-related mortality rates remained at the same level. However, by 2017 that amount has tripled.
A similar situation is visible across Europe. United Kingdom observed a quadruple increase in opioid prescriptions between 2000 and 2010, with morphine being the most frequent prescription (Zin, Chen, & Knaggs, 2014). Opioid use in Germany increased by 15 percent in that same time period (Schubert, Ihle, & Sabatowski, 2013). Between 2004 and 2017, the use of strong opioids more than doubled in France (Chenaf et al., 2019).
The United States of America has been touched the most by this crisis. The rapid increase in prescription opioid use resulted in higher mortality rates. Between 1999 and 2007 this number has tripled (Dhalla, Persaud, & Juurlink, 2011). During that time, studies have shown a five-time increase in drug treatment admissions (Set, n.d.), while overdose deaths increased from 3000 to 12000 (Network, 2012). Currently, death by overdose is the second leading cause of unintentional death in the United States (for Disease Control, Prevention, et al., 2013). The marked increase in opioid prescriptions over recent decades can have its roots in several factors. Firstly, the expansion of opioid use into chronic non-cancer pain treatment increased the total number of prescriptions (Dowell, Haegerich, & Chou, 2016). Pharmaceutical marketing, particularly in the United States, also played a crucial role, with various campaigns promoting opioids for common use (Van Zee, 2009).
Although the issue of opioid-related mortality rates in the Netherlands remains lower than in the United States, it represents a worrying trend that should not be overlooked. Opioid use, especially prolonged or misused, is associated with a range of negative outcomes, including an increased risk of dependence, overdose, and death (Pergolizzi Jr, Raffa, & Rosenblatt, 2020). Beyond the well-documented risks of overdose and dependence, prescription opioid use is associated with a range of social and psychological consequences. They include depression (Scherrer et al., 2016), sleep disturbances (Webster, Choi, Desai, Webster, & Grant, 2008), or cognitive decline such as dementia (Sun, Chen, Wu, & Zhang, 2023). Identifying and understanding these associations is crucial in order to gain awareness of the possible social and psychological risks associated with opioid use.

1.2 Research Question

This paper will build on previous work by Ramya Tumkur Rameshchandra (Tumkur Rameshchandra, 2024), which established baseline associations between opioid prescriptions and social or psychological outcomes in Dutch general practice data. This time, it will focus on an application

of an alternative machine learning (ML) principle called Minimum Description Length (MDL). MDL-based method aims to extract high-quality patterns in the data (Galbrun, 2022), which is crucial for this study. It is a principle used for model selection and is commonly applied in data mining and machine learning applications. MDL-based approach has proven to be a promising tool in the recent study conducted by Lincen Yang, where it was able to successfully balance the complexity and fit of the model (L. Yang & van Leeuwen, 2024). This advantage is particularly valuable in dealing with noisy and complex data, which is the case in medical datasets like ELAN, where numerous factors and entries occur.

The main research question of this thesis is as follows:
RQ: To what extent does an MDL-based rule-mining approach improve over baseline association rule mining in the analysis of social and psychological aspects of opioid prescription data?

To thoroughly investigate this topic, the following sub-questions can be answered during the study:

- SQ1: How do the results from MDL method compare to those obtained during ARM in terms of understanding the social and psychological effects of prescription opioids?

- SQ2: What are the advantages and limitations of MDL-based rule mining in identifying patterns within the ELAN dataset?

1.3 CRISP-DM Framework

This research will largely follow the CRISP-DM framework. The Cross Industry Standard Process for Data Mining (CRISP-DM) is a widely recognized methodology for data mining projects. Originally developed in the late 1990s (Wirth & Hipp, 2000), CRISP-DM provides a repeatable framework that supports various research analyses. Its iterative nature allows researchers to make adjustments when needed and as new insights arise.
The CRISP-DM framework is based on six interconnected phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. They can be seen in Figure 1. In the context of this study, which examines associations between social and psychological factors in opioid prescription patterns using advanced rule mining, these phases are adapted as follows.

1. **Business Understanding** - Here, the focus has been put on the literature review on the current state-of-the-art research in the medical field regarding the topic of opioids. Moreover, a previous study on the Dutch GP dataset has been studied to understand its results and limitations. MDL-based rule mining methods, particularly the Truly Unordered Probabilistic Rule Sets (TURS), were examined to see the additional value they can provide for analysis of medical datasets. The purpose of this step is to understand the current state of research and notice its gaps.

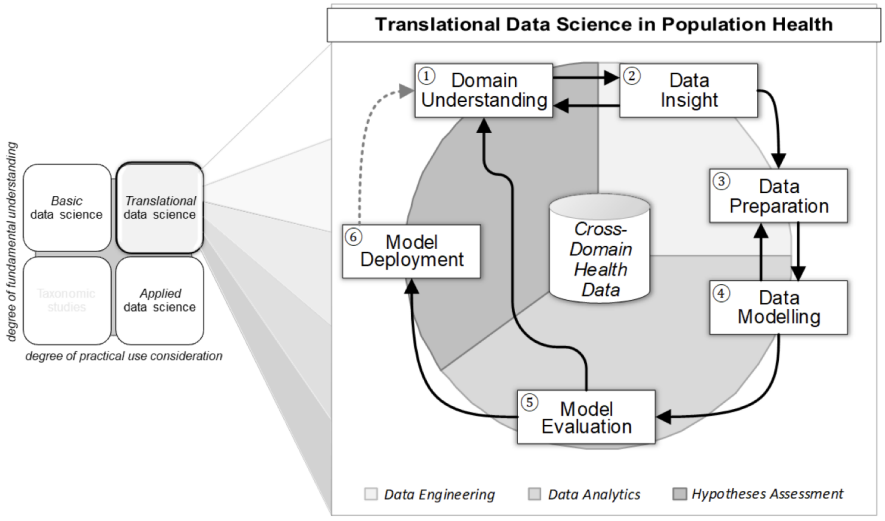

Figure 1: CRISP-DM Framework in a Population Health contexts based on (Spruit, 2022).

2. **Data Understanding** - Dataset has been collected and studied. In this project, the dataset consists of three linked components: patient records, clinical episodes, and medication prescriptions. The files were analysed to gain an understanding of the contents of the dataset. Data was assessed in terms of null values, and relevant variables were identified.

3. **Data Preparation** - Crucial step that is the foundation for further reliable modelling. For this step, data was merged, filtered, and prepared for analysis. Special attention was paid to extracting needed features and transforming them into a proper format for the TURS model.

4. **Modelling** - TURS algorithm has been applied with the MDL principle to identify probabilistic rule sets.

5. **Evaluation** - The results were evaluated against the research objectives, and the research questions were discussed in depth.

6. **Deployment** - Traditional CRISP-DM involves implementing the results in a real world environment. While this thesis does not do this exactly, the step is instead addressed through reflection on the work and discussion about the highlights and drawbacks of the results.

1.4 Outline

This thesis is organized into different sections. The abbreviation section lists concepts that show up frequently in this paper and are important for the research. Chapter 1 introduces the Research Question, its motivation, and describes the methodology framework. Chapter 2 reviews the relevant literature, covering opioid-related research in medical and public health contexts, traditional association rule mining approaches, and recent developments in MDL-based pattern discovery. Chapter 3 describes the dataset and its preprocessing steps. Chapter 4 describes the methodology for the analysis. Chapter 5 presents the results obtained during the study. Chapter 6 discusses the

findings obtained in the MDL-based approach and compares them with those from the baseline association rule mining. Chapter 7 summarizes the main conclusions and reflects on the strengths and limitations of the study. Chapter 8 discusses potential directions for further research.

2 Background and Related Work

The literature review was performed using the Snowball method (Wohlin, 2014). It is an approach in which relevant sources are tracked down by examining the reference lists and bibliographies of other papers.

2.1 Opioids and Opioid Use Disorder

Opioids are substances commonly used for pain treatment. They consist of compounds that are able to act on the opioid receptors of the body. There exist three known and documented receptors - mu (μ), delta (δ), and kappa (κ) receptors (Al-Hasani & Bruchas, 2011). The term 'opioid' refers to natural, semi-synthetic, and synthetic substances, while the term 'opiate' is used to describe only natural substances. Opiates such as morphine and codeine are derived directly from the opium poppy. Semi-synthetic variants include substances such as oxycodone and hydrocodone. Synthetic opioids, such as fentanyl, are created in laboratories.
Although opioids can be highly effective in alleviating pain, their use carries a substantial risk of dependence and addiction. This, in turn, can result in various health and social challenges. When opioid use progresses to a pattern of compulsive consumption, it is classified as Opioid Use Disorder. Opioid Use Disorder (OUD) refers to the chronic condition characterized by persistent use of opioids despite its negative consequences, the development of tolerance, and withdrawal symptoms after cutting the use of opioid substances (Taylor & Samet, 2022). Risk factors for OUD include current or past opioid abuse, young age of the person, or social environment that encourages the misuse of such substances (Webster, 2017). By 2020, the United States estimated around 3 million OUD occurrences among its citizens (Dydyk, Jain, & Gupta, 2024).

2.2 Machine Learning in Healthcare and Opioid Research

Machine learning, in its various forms, contributed to the field of healthcare in a lot of ways. For example, the Support Vector Machine (SVM) has been used for the diagnosis of heart failure, scoring an accuracy of 74,44% (G. Yang et al., 2010). The decision tree classifier scored an accuracy of 86,5% (Aljaaf et al., 2015). ML has also been used in decision support in infectious diseases, the detection of diabetic retinopathy, or the health emergency management (G. Yang et al., 2010) (Gulshan et al., 2016) (Lu, Christie, Nguyen, Freeman, & Hsu, 2022). ML methods are also used in opioid research. One study evaluated the risk of opioid overdose among opioid prescription patients (Lo-Ciganic et al., 2019). It utilized least absolute shrinkage and selection operator-type regression (LASSO), gradient boost machine (GBM), deep neural networks (DNN) and random forest (RF) which aimed to predict chances of overdose risk after initial prescription. GBM ad DNN models outperformed the rest, achieving high prediction scores and grouping patients into three different risk groups. Another study used electronic health records (EHR) to study the tendencies of opioid-dependent patients to predict risks of overdose and opioid dependence (Ellis, Wang, Genes,

& Ma'ayan, 2019). Through the Random Forest method, it was concluded that opioid-dependent patients showed higher amounts of white blood cell (WBC) counts, respiratory issues, higher rates of psychiatric conditions, as well as showed signs of malnutrition. A study in the United States focused on the application of ML in order to better predict opioid misuse among adolescents (Han, Lee, & Seo, 2020). Data, concerning youth between 12 and 17 years from 2015 to 2017, was analysed with artificial neural networks (ANN), distribution random forest, and gradient boosting. The results of all methods were similar and showed that 3,7% (1521) of the studied adolescents have misused opioids before. Another study focused on finding an algorithm for reliable early OUD detection (Segal et al., 2020). It used data from 550000 patient records and applied Word2Vec alongside Gradient Boosting trees algorithm. The new model allowed for a 14.4 month earlier diagnosis of OUD.
As can be seen, there exists numerous research on the topic of opioid misuse and OUD. However, there is a gap in understanding the psychological and social effects of opioid prescription. In her master thesis, Ramya (Tumkur Rameshchandra, 2024) studied these correlations in Dutch medical data, implementing baseline machine learning methods such as Apriori and FP Growth. Through her study, she was able to identify interesting patterns that will be discussed in more detail in Section 5.1. This paper will continue her work through the implementation of an advanced MDL-based ML technique, as it has not yet been used in the opioid domain.

2.3 Minimum Description Length Principle

The Minimum Description Length (MDL) is a principle that is the core foundation of the TURS algorithm. MDL is a method of model selection grounded in information theory and statistical inference. Over the decades, MDL has evolved into a powerful tool and has been applied in various domains, such as malware detection or text summarization (Asadi & Varadharajan, 2019) (Vanetik & Litvak, 2018). It was originally introduced by Rissanen (Rissanen, 1978) as a formalization of Occam's razor, which is a principle that states:

> All else being equal, simpler explanations are preferable to more complex ones (Baker, 2022)

This means that when two potential explanations fit the data equally well, we prefer the simpler one. MDL applies this idea through the number of bits required to describe the model and the data when using that model. It aims to balance between oversimplifying and overcomplicating the model. The reason for this is that very simple models (short L(M)) cannot capture enough of the structure in the data, which in turn can lead to missing important patterns. On the other hand, overly complex models result in large L(M) as the model description is long. MDL principle searches for the model that will minimize the total number of bits. Mathematically, it is described as follows:

$$\text{MDL}(M) = L(M) + L(D \mid M)$$

where
$L(M)$ - description length, in bits
$L(D \mid M)$ - length of the encoded data D

Here, $L(M)$ is responsible for capturing the complexity of the model, while $L(D \mid M)$ describes the fit of the model to the data. The role of the MDL principle is to strike a balance between these two components (Galbrun, 2022). This means that patterns are selected when they contribute to explaining the data more concisely, not just because they occur frequently. For example, the KRIMP algorithm (Dutch for 'to shrink') uses MDL by maintaining a code table that contains both individual items and larger patterns, assigning shorter codes to those patterns that contribute most to compression (van Leeuwen, Vreeken, & Siebes, 2006). During model building, KRIMP iteratively replaces less useful patterns with ones that lead to better compression. In this thesis, MDL is applied in the Truly Unordered Probabilistic Rule Sets (TURS) approach, which is another example of the MDL-based algorithm that will be introduced in the following section.

2.4 Truly Unordered Probabilistic Rule Sets

Rule-based models are characterized by their interpretability. Due to that, they have been gaining more attention in areas were research is related to sensitive real-world scenarios, such as healthcare or justice systems (Rudin, 2019). However, they have their limitations. Firstly, rule set learning imposes an order - rules are applied sequentially, and the first matching rule dictates the outcome. Secondly, rules often overlap, and this problem is not being well handled, which leads to lower interpretability and overall performance. To address these issues, Truly Unordered Probabilistic Rule Sets approach was introduced by Lincen Yang and Matthijs van Leeuwen (L. Yang & van Leeuwen, 2024). TURS addresses both of those problems by removing rule ordering and incorporating probabilistic reasoning. Probabilistic rules have a form of :

$$\textbf{If } X \text{ meets certain conditions, } \textbf{then } P(Y) = \hat{P}(Y),$$

where X is the feature variables, Y is the target variable, and $\hat{P}$ is the associated class probability estimator. (L. Yang & van Leeuwen, 2024)

While one rule describes only a small subset of the whole dataset, rule-based models aim to gather a certain number of such rules and form one global predictive model, which then can be easily read by both data scientists, as well as people who are not related to this field. This is due to the fact that outcome rules come in an easily understandable form.
To adress the issue of implicit rule orders, TURS allows rules to overlap only if they have similar probability. Therefore, it does not matter which one is chosen, and thus each of them becomes 'independent'. TURS minimizes the total description length of the dataset by incrementally adding rules that contribute most to compression. Unlike other baseline machine learning models, TURS does not aim to find all frequent rules. Instead, it identifies a compact rule set that explains the data efficiently. The crucial differences between the TURS and baseline ARM methods are described in Table 1.

2.4.1 How TURS works

This section describes how the algorithm works, based on its pseudocode and the explanation in (L. Yang & van Leeuwen, 2024).
At the start, data is encoded with a use of `OneHotEncoder`. The algorithm begins with an empty rule set M and an initial 'Empty Rule'. At each iteration, two search beams of width W are kept.

Table 1: Comparison between TURS, Apriori, and FP-Growth algorithms.

	TURS	**Apriori**	**FP-Growth**
Task	Multi-class classification using probabilistic unordered rule sets.	Frequent itemset mining for association rule generation.	Frequent itemset mining using FP-tree compression.
Principle	Minimum Description Length (MDL) for model selection, beam search for candidate exploration.	Breadth-first search over item combinations	Depth-first pattern growth
Rule ordering	Truly unordered	Ordered	Ordered
Output	Compact probabilistic rule sets	Deterministic association rules (antecedent $\rightarrow$ consequent)	Same as Apriori
Interpretability	High — rules are short, probabilistic, and overlap problem is resolved.	Moderate — rules can be numerous and overlapping	Moderate — rules can be numerous and overlapping

The primary beam tracks the top-W candidate rule expansions based on a learning speed score $r(\cdot)$. The second auxiliary beam keeps the top-W rule expansions ranked by a complementary score $R(\cdot)$, which ignores any coverage overlap with the existing rule set. Candidate rules are created by adding one new literal to each base rule. The numerical and categorical variables are being split. Candidates are grouped by coverage for the primary beam and by residual coverage (excluding instances already covered by M) for the auxiliary beam. An MDL-based test is applied to each group, and the best rule in each group is kept.
The process repeats until a stopping condition is met. This happens when, for K_{stop} iterations, neither beam finds better scoring rules than in the previous iteration. If this condition is not met, both beams are reduced to width W by clustering on coverage and keeping the top-W rules per cluster. The resulting rule sets are the base for the next iteration. Once the stopping condition is satisfied, the algorithm selects and returns the candidate rule from all candidates with the highest score $r(\cdot)$ and then considers whether to include it in the rule set M.

3 Dataset

3.1 Data Access

The dataset used in this study comes from the Extramural Leiden Academic Network (ELAN). It contains records from over 100 general practitioners (GP) in Leiden and The Hague. To gain access to the ELAN dataset, a formal request has been sent to LUMC faculty, and work was possible through the LUMC network that enabled access to the PHEG departmental drive.

3.2 Dataset Description

3.2.1 Medical Coding Systems

The dataset uses medical coding systems that represent information about the patient in a consistent way in medical records. In this paper, two coding systems are present. The International Classification of Primary Care (ICPC) is a classification system for primary care analyses. Established in 1984 by the World Organization of Family Doctors (WONCA), it is now widely used to register diagnoses and health problems in standardized categories (Bentsen, 1986). Anatomical Therapeutic Chemical Classification System (ATC), supported by the World Health Organization (WHO), classifies medications according to the organ or system they act on.

3.2.2 Dataset Overview

Data used in the analysis included five separate files:

1. **Two Patients files**: Contain demographic and registration information about the patients, such as gender, birth year, marital status, education, and registration dates.

2. **Episodes file**: Contain clinical episodes, their ICPC codes, start and end dates, and episode types.

3. **Two Medications files**: Contains prescribed medications, including ATC codes, dosage information, and prescription dates.

Patient files and Medication files were joined respectively, as they related to the same information. A more detailed overview of the data per file can be seen in Table 2, 3, and 4.

3.3 Data Filtering

The analysis focuses on exploring the relationship between psychological and social health aspects of opioid prescription data. Inclusion criteria included:

- Psychological and social health aspects described by ICPC codes starting with 'P' and 'Z', respectively.
- Above ICPC codes had to be diagnosed between 01/01/2010 and 31/12/2019.
- Prescription of opioid medications were marked by ATC codes starting with 'N01AH' or 'N02A'.

Therefore, initial data filtering is based on these three indicators.
Firstly, episodes with ICPC codes beginning with "P" or "Z" between the years 2010 and 2019 were extracted. Patients and medications files were then filtered to include only those linked to the selected episodes. All files were merged on Patient ID field (PATNR) which is the only field that connected all files. EpisodeID proved to have no correlation between Episode and Medication files. Only parent ICPC codes were considered, which means that only the first three characters were considered (for example, all occurrences of Z16.03 were reduced to Z16)
As can be seen in Table 2, 3, and 4, many columns have large amount of missing values, which severely affects further research. Variables such as country, martial status or age could not be included in the research, as imputation in those cases would result in unreliable results. Therefore, gender has been assessed as a promising feature.
Each file was checked for duplicate entries, which were removed. The ICPC and ATC codes were then aggregated by patient.

3.4 Data Analysis

Data analysis was performed using the `matplotlib`, `numpy`, and `seaborn` libraries in Python. This exploratory stage provided key insights into the dataset and patient characteristics, as illustrated in the figures below. Between 2010 and 2019, a total of 153,919 patients were prescribed opioids. Majority - 62.32% - were women while the remaining 37.68% were men.
Figure 2 represents a bar plot showing the distribution of ATC codes among patients. It can be seen that a small subset of opioid prescriptions dominates the dataset, while the rest is rarely prescribed. By far, the most frequent medication was Tramadol (N02AX02), which was prescribed 54454 times. Next in line were Fentanyl (N02AA05) and Oxycodone (N02AJ06).
Figure 3 shows the same distribution, however here it is also divided by gender. It should be noted that in all visible cases women are the majority of the patients with prescribed opioids. It is especially visible in the case of N02AX02, where only around 20000 men were prescribed the

Variable	Translation	Non-null Count
PATNR	Patient Number	313776
PRAKNR	Practice Number	313776
Woonverband	Residential number	292757
dWoonverbandsoort	Type of residence	17324
dWoonverbandpositie	Position in patient's living environment	19650
dPostcodecijfers	Postcode	312105
iGeboortejaar	Year of birth	313776
iOverlijdensjaar	Year of death	757
dGeslacht	Gender	313776
Thuisland	Country	8173
dBurgerlijkeStaat	Marital status	15430
Beroep	Profession	1557
dOpleiding	Education	59
dInschrijfdatum	Date of registration with healthcare provide	309696
dUitschrijfdatum	Date of deregistration with healthcare provider	59273
dRedenVertrek	Reason for deregistration	55196

Table 2: Non-null count of variables in Patient Data.

Variable	Translation	Non-null Count
PATNR	Patient Number	1003117
PRAKNR	Practice Number	1003117
EpisodeID	Episode Number	1003117
dBegindatum	Start date of episode	1003117
dEinddatum	End date of episode	166780
dMutatiedatum	Date of last change in episode registration	900466
dICPC	ICPC code registered during episode	1003117
dEpisodetype	Indicator for the type of episode	992696
dActief	Episode activity	900466
dAttentie	Indicator for the attention value of the episode	55257

Table 3: Non-null count of variables in Episode Data.

Variable	**Translation**	**Non-null Count**
PATNR	Patient Number	1177433
PRAKNR	Practice Number	1177433
EpisodeID	Episode Number	1119492
dVoorschrijfdatum	Prescription date	1177433
dEinddatum	Date until which medicine was prescribed	1076319
dStopdatum	Date on which medication was stopped	8718
Etiketnaam	Name of medicine	1135055
dPRK	KNMP-Prescription code	1067096
dGPK	KNMP-Generic product code	1067325
dATC	ATC Code	1177433
dChronisch	Chronic medication indicator	51096
dDuur	Prescription length in days	51466
dIteraties	Number of permitted repeat prescriptions	2251
dHoeveelheid	Quantity	1177075
Dosiscode	Dosage code	1158590
dSterkte	Strength of drug	1007153
dToedieningomschrijving	Route of medication administration	802138
dVoorschriftICPC	ICPC code of diagnosis for which medication was prescribed	295500
dEpisodeICPC	ICPC code of episode	437826
dSpecialisme	Medical speciality of the prescriber	1044430

Table 4: Non-null count of variables in Medication Data.

substance, in contrast to around 34000 women.
Figure 4 shows a respective representation for ICPC codes. The most frequent diagnosis concerns insomnia and other sleep disorders (P06), occurring more than 20000 times. Tobacco abuse (P17) and feeling of anxiety (P01) take second and third place, respectively.
Figure 5 represents the gender division of ICPC diagnoses. Here, women are also a majority in every ICPC occurrence.
Frequency of prescriptions over time can be seen in Figure 6. It is clear that over the years there was an increasing trend in opioid prescription, with a value around 90000 in 2010, and more than 130000 in 2018. However, the number slightly decreased in 2019 to around 120000 opioid prescriptions.
Figure 7 represents the age distribution of the patients at the time of the prescription of the medicine. Age has been determined by subtracting the birth year from the prescription date. It is clear that the highest prescription rate occurs between 50 and 60 years, slowly dropping over time. Moreover, it is worth noting that while it may not seem significant, the young age group between 20 and 30 years old has been prescribed opioids around 45000 during the given period of time. It seems to confirm the fact that opioid usage concerns almost all age groups.

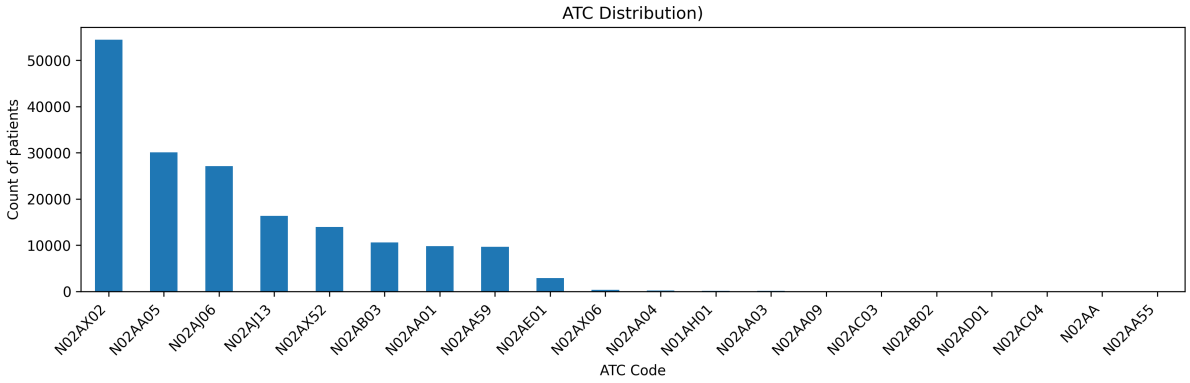

Figure 2: Distribution of ATC codes among patients.

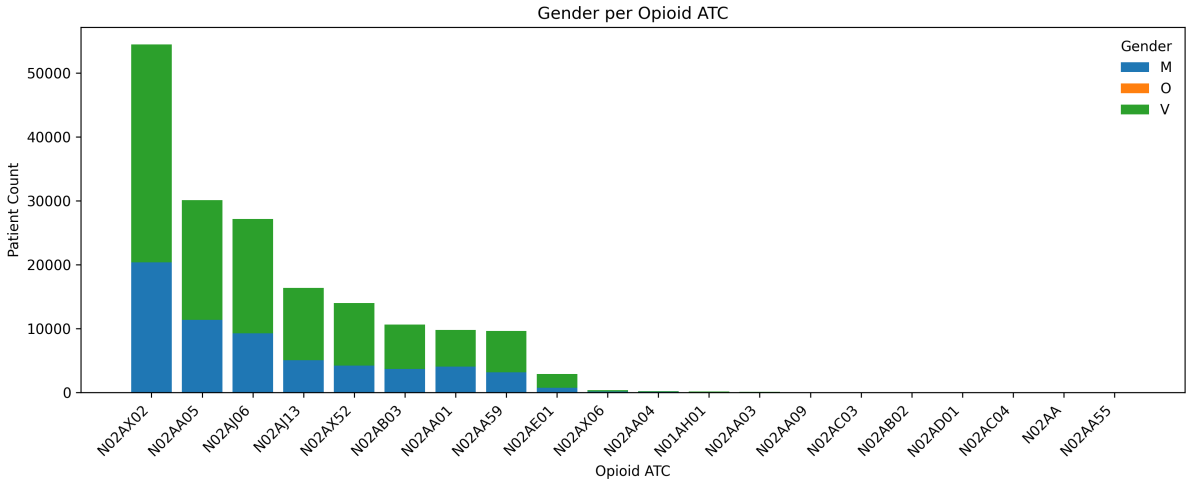

Figure 3: Top 20 ATC codes among patients with gender distribution.

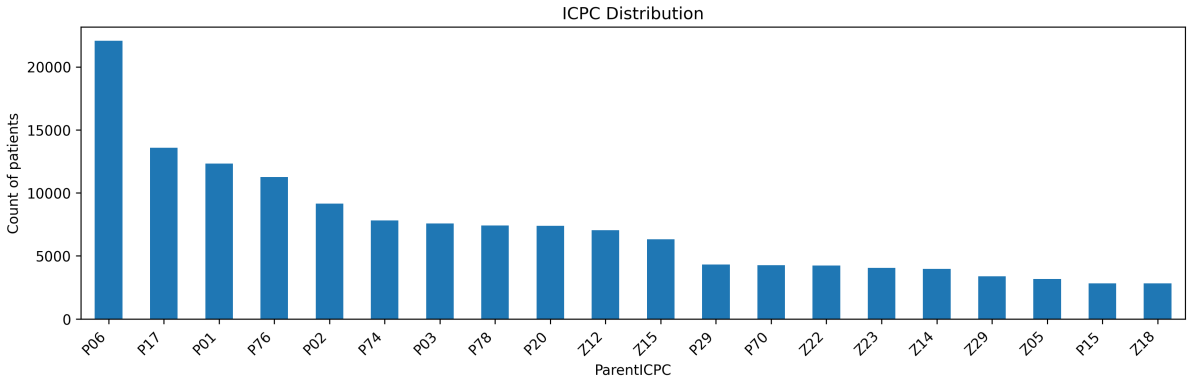

Figure 4: Distribution of ICPC codes among patients.

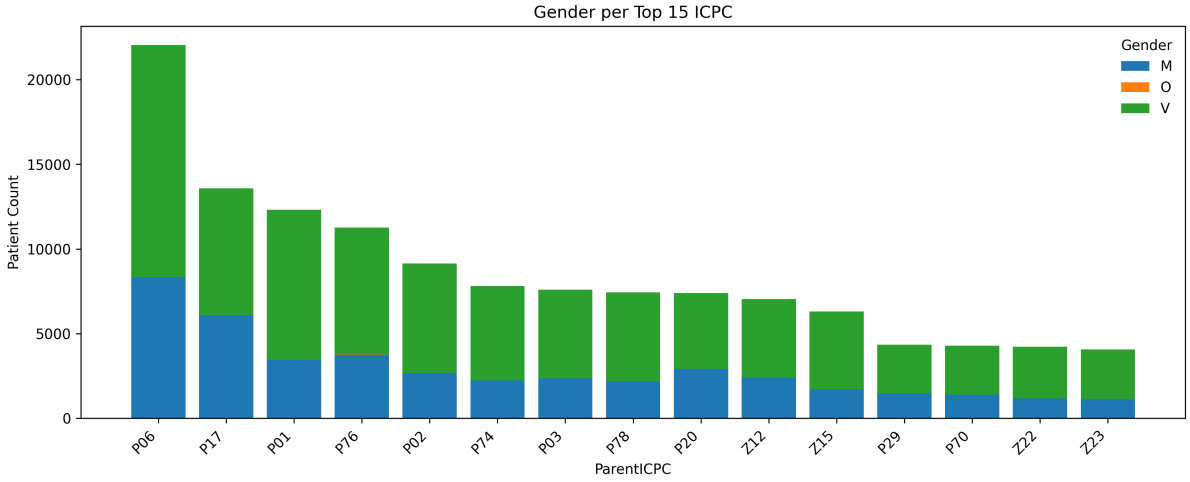

Figure 5: Top 15 ICPC codes among patients with gender distribution.

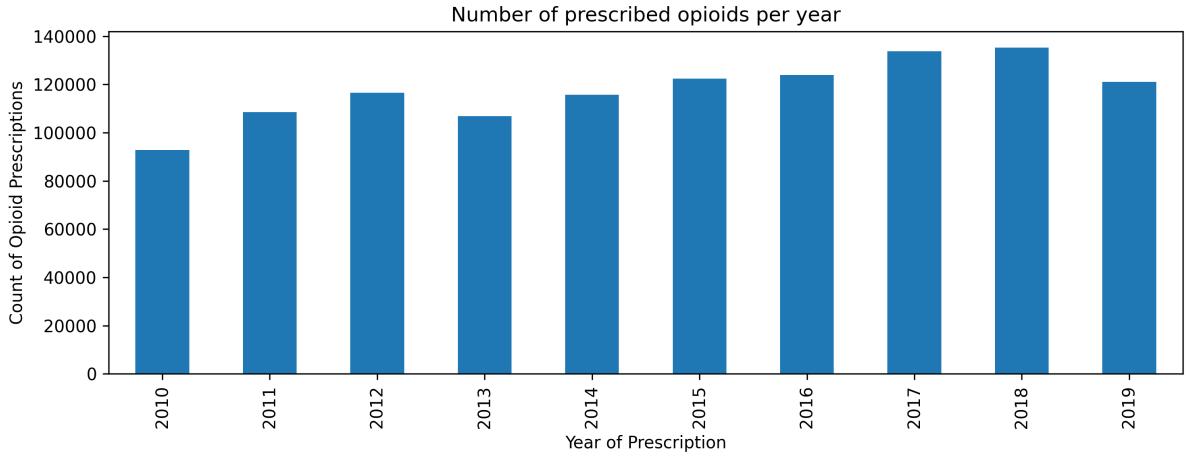

Figure 6: Number of prescribed opioids per year between 2010 and 2019.

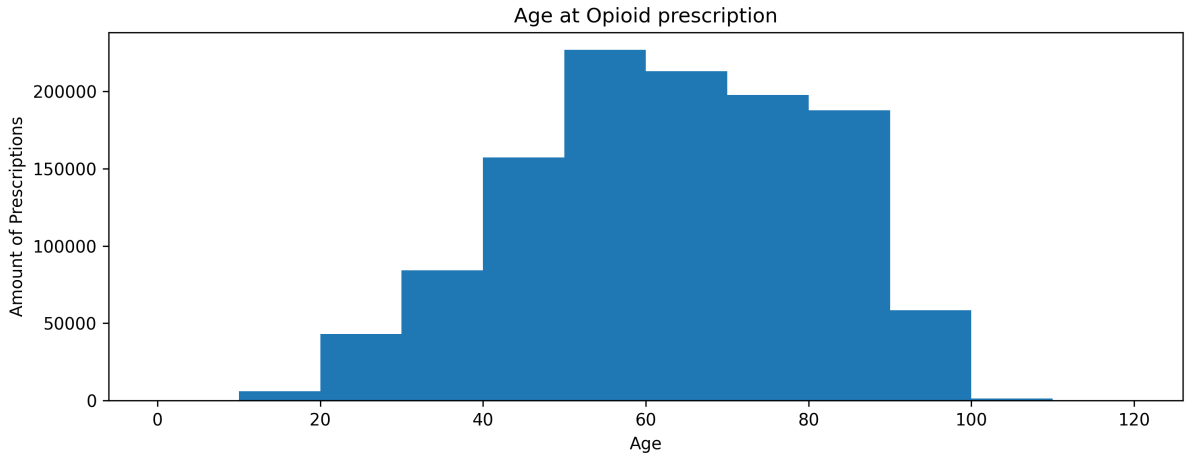

Figure 7: Age distribution of the patients at the time of the prescription of the medicine.

3.5 Feature and target construction

Multiple data transformation strategies were explored to effectively prepare the dataset for analysis using the MDL-based TURS algorithm. Each strategy attempted to maintain data granularity, minimize sparsity, and align with the input requirements of TURS. Therefore, it was crucial to pass properly formatted dataset to derive meaningful rules. The main approaches considered in this research were as follows:

1. **Wide Format with One-Hot Encoding**: Each ICPC and ATC code was converted into its own binary feature column (1 = present, NaN = absent). However, this resulted in a highly sparse matrix that resulted in no rules found.

2. **Compact Patient-wise Format**: An attempt was made to structure the data set so that each row corresponded to a single patient, with a column for gender, a list of ICPC codes, and a list of ATC codes. Although semantically intuitive, this format caused an important issue. The first concern was losing granularity, as TURS was not able to encode a list of codes in one column as separate features. Instead, it encoded each list as a whole entity and thus was excluded from further analysis.

3. **Row-wise Long Format**: The most promising method transformed the dataset into a long format, where each row represented a single feature-target pair. For example, a patient with three ICPC codes and one ATC code would be represented by three rows. This preserved granularity and eliminated false negatives from NaN or 0 values. This method means that there will be more than one row per patient. This method was chosen as the most promising option and was considered further.

An additional challenge was that TURS requires the data to consist of a target column that corresponds to the part of the rule "THEN". The target column needs to be placed as the last column in the dataset, therefore a proper preprocessing was done. Two different target columns were chosen to align with the research question, and thus the dataset was prepared in two different ways. The first target column was based on ICPC codes and the second column was based on ATC

codes. In this way, it was possible to obtain a more complete understanding of the patterns and correlations between opioid use and social and psychological diagnoses.
Dimensionality reduction was applied across all approaches in order to ensure the effectiveness of the model. It was done by setting a threshold and excluding all ICPC and ATC codes that fell below that value from the TURS analysis. This reduced the rule space, improved runtime, and did not harm the output of the model since the results prior to this reduction did not consider those codes in any way. The most frequent ACT codes that were included in further research are ordered by frequency and can be seen in Table 5 and 6. The most frequent P/Z ICPC codes that were diagnosed in further research are ordered by frequency and listed below.

ATC code	**Medication**
N02AX02	Tramadol
N02AB03	Fentanyl
N02AA05	Oxycodone
N02AA01	Morphine
N02AJ06	Codeine and Paracetamol
N02AJ13	Tramadol and Paracetamol
N02AA59	Codeine, combinations excluding psycholeptics
N02AX52	Tramadol combinations
N02AE01	Buprenorphine

Table 5: Most frequent ACT codes, exceeding the set threshold.

ICPC code	Diagnosis
P06	Insomnia/other sleep disorder
P17	Tobacco abuse
P01	Feeling anxious/nervous/tense
P03	Feeling down/depressed
P78	Overvoltage
P20	Memory/concentration/orientation disorders
Z12	Relationship problem with partner
Z15	Loss/death of partner
P70	Dementia
P29	Other psychological symp- toms/complaints
Z05	Problem with work situation
P15	Chronic alcohol abuse
Z18	Problem with child's illness
P71	Other organic psychosis(s)
P21	Attention-deficit/hyperactivity dis- order
Z16	Relationship problem with child
P76	Depression
P02	Crisis/transient stress response
P74	Anxiety disorder/anxiety state
P99	Other mental disorders
Z25	Problem due to violence
P22	Other child behavior concerns
P18	Drug abuse
P04	Irritable/angry feeling/behavior
Z03	Housing/neighborhood problem
P19	Substance abuse
P12	Enuresis [ex. U04]
P24	Specific learning difficulty
Z22	Problem with illness of par- ents/family
Z20	Relationship problem with par- ents/family
Z23	Loss/death of parents/family
Z14	Problem with partner's illness
Z29	Other social problem

Table 6: Most frequent ICPC codes, exceeding the set threshold.

4 Methodology

4.1 Baseline ML with Apriori and FP-Growth

A prior study conducted by Ramya Tumkur Rameshchandra (Tumkur Rameshchandra, 2024) examined the same social and psychological correlates of prescription opioid use on ELAN data from 2010 to 2019. The dataset consisted of a total of 313,776 patients, where opioid prescriptions was identified using ATC codes in the N01AH* and N02A* categories. Diagnoses were identified by ICPC P (psychological) and Z (social) codes, and data preparation was made on the same basis as the one in this paper, excluding final data transformation step for feeding the model. Here, tuples with gender, ICPC codes and ATC codes per patient were used as an input for the models.
The analysis applied baseline association rule mining (ARM) techniques, using the Apriori and FP-growth algorithms to find patterns between patients' gender, diagnoses, and prescribed opioids. Rules were checked through the following measures:

- **Support** — How frequently an itemset appears in the dataset:

$$\text{support}(A \to B) = \frac{\text{number of transactions containing } (A \cup B)}{\text{number of total transactions}}$$

  Used in Apriori and FP Growth to filter itemsets that occur often enough to be statistically relevant.

- **Confidence** — Reliability of the rule:

$$\text{conf}(A \to B) == \frac{\text{supp}(A \cap B)}{\text{supp}(A)}$$

  High confidence means B frequently follows A.

- **Lift** — Measures how much more likely B is given A compared to B occurring independently:

$$\text{lift}(A \to B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A)\,\text{supp}(B)}$$

  Lift > 1 indicates a positive association; Lift < 1 indicates a negative association.

- **Conviction** — Measures the degree of dependence, considering the frequency of A without B:

$$\text{conviction}(A \to B) = \frac{1 - \text{support}(B)}{1 - \text{confidence}(A \to B)}$$

  Higher conviction indicates stronger dependence.

- **Leverage** — Measures the difference between the observed co-occurrence of A and B and what would be expected if they were independent:

$$\text{leverage}(A \to B) = \text{support}(A \cup B) - \text{support}(A) \times \text{support}(B)$$

  Values close to zero indicate independence; positive values mean positive correlation.

4.2 TURS set up

TURS iteratively searches for rules that compress and describe the dataset best. The implementation had the following parameters:

`max_num_rules`: **500** Maximum number of rules allowed in the model. It is a safe high threshold to ensure all important rules will be included.

`max_grow_iter`: **500** Maximum iterations allowed during rule search.

`beam_width`: **10** Width of the search beam for rule candidates. It encourages diversity in candidate rules without overwhelming the memory usage.

`candidate_cut_points`: **20** for numeric features

The analysis is performed through five-fold stratified cross-validation. Datasets that were provided to the model have been described in section 3.5. TURS requires the last column of the data to be a target label. Thus, one of the dataset had ATC codes as the target column, while the second had ICPC codes as the target.
Several metrics were used to show the performance of the model. `Probability` metric assesses how confident the rule is in predicting a class. `Coverage` outputs the amount of instances when a certain rule was seen in data. `Train and test probability difference` shows how stable the probability is between the train and test data. Lower score implies more stable rules. `ROCAUC score for test and train data` measures model's ability for generalization - high scores for both components mean that TURS generalizes well. If `Generalization gap`, measured by the difference of those two scores, will be high, it will imply overfitting of the model. `Average rule length` puts shorter rules in favor, as TURS aims to achieve compactness through the MDL principle. Another important metric is `Overlap`. It will look at those instances that were covered by more than one rule. For good TURS performance, this score should be minimal.

4.3 Comparative rule analysis

Evaluation metrics for baseline ML approaches and TURS are not directly comparable. TURS provides probability and coverage for each rule, while the Apriori and FP-Growth rules are described by support, confidence, lift, leverage, and conviction. For this reason, the alignment was based primarily on the content of the rules, that is the overlap between the antecedent conditions and the target outcomes. The overlap was measured using the Jaccard similarity of the rules. For two sets A and B representing the items of a TURS rule and a baseline rule, the Jaccard similarity is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad 0 \leq J(A, B) \leq 1.$$

Based on this calculation, each rule combination, if it existed, was assigned to one of the three categories according to the relevant Jaccard score threshold. The thresholds were chosen to balance strictness and inclusion, as well as take into account the relatively low average amount of features in each rule. `Exact` match describes a perfect overlap, that is all parts of the rule of one method are present in the other. `Close` alignment makes sure that rules are considered similar when two-thirds

of their conditions overlap. A lower bound of 0.4 defines the `Partial` alignment, while values below 0.4 are classified as `Novel`:

$$\text{Alignment}(A,B) = \begin{cases} \textbf{Exact} & \text{if} \quad J(A,B) = 1 \\ \textbf{Close} & \text{if} \quad 0.65 \leq J(A,B) < 1 \\ \textbf{Partial} & \text{if} \quad 0.40 \leq J(A,B) < 0.65 \\ \textbf{Novel} & \text{if} \quad J(A,B) < 0.40 \end{cases}$$

To evaluate to what extent an MDL-based TURS improves over the baseline association rule mining, the rules retrieved by TURS were aligned with the 133 rules previously obtained using Apriori and FP-Growth. The complete overview of the baseline rules can be seen in Table 9 in (Tumkur Rameshchandra, 2024).
Each TURS rule is translated to English, and matching IDs from previous research are provided. In case no relevant matching rules were found, a `No matching baseline rule` message is displayed.

5 Results

5.1 Baseline study

The application of Apriori and FP-growth resulted in over 130 rules. In the end, the ones that were considered important are listed below:

- Higher opioid perscriptions among women
- Fentanyl and morphine associated with dementia
- Fentanyl, tramadol, and morphine correlated with memory, concentration, and orientation disorder
- Correlation between many opioids and sleep disorders
- Tramadol and oxycodone associated with depression
- Tramadol and oxycodone prescribed to patients who reported the loss or death of a partner
- Fentanyl, morphine, and codeine negatively associated with tobacco abuse
- Chronic alcohol and substance abuse occurred mainly among men
- Women associated with having problems with child's illness
- Attention deficit and hyperactivity disorder associated more with men

The issue with this approach was the noisy and overlapping output. Apriori and FP-growth select rules independently from each other. Thus, they do not optimize them globally, which causes redundancy. In an attempt to overcome this problem, TURS is implemented on the same dataset. Additional exploratory analyses were performed, including multiple correspondence analysis (MCA), autoencoders, and temporal long short-term memory (T-LSTM) models with an aim to create patient clusters and subtyping. However, these methods provided clusters that were too large and not suitable for clinical interpretation, thus they were not considered further.

5.2 Truly Unordered Probabilistic Rule Sets

Analysis of the first dataset (ICPC as a target) resulted in 12 rules. The second dataset (ACT as a target) resulted in 20 rules. The results can be seen in Table 7, where each rule is assigned its own ID number. Immediately, it is noticeable that the predicted probabilities for all rules are relatively low. This is expected as TURS is designed for multiclass classification. This means that the overall probability (that sums up to 1) must be distributed across multiple possible classes. Consequently, even the most confident rule will rarely reach a high probability score.
Table 8 provides model performance metrics. There are several correlations that can be derived from the rulesets:

- Prescription opioids (i.e Buprenorphine, Tramadol) are often correlated with insomnia or other sleep disorders (1.3, 1.7)
- Strong correlation between being a man, being prescribed Oxycodone, and struggling with insomnia (1.6)
- Corelation between being a man, dementia, and being prescribed Tramadol (2.2)
- Corelation between Tramadol prescription and relationship problems with a partner or loss of a partner (2.7, 2.8, 2.12)
- Strong corelation between feeling depressed and Tramadol prescription (2.19)
- Corelation between being a man, tobacco abuse and Tramadol prescription (2.13)
- Corelation between memory disorders and Tramadol prescription (2.1, 2.2, 2.3, 2.5)

Several rules are complementary in terms of gender, and when put together, they create a more informative correlation. Rules 2.3 and 2.5 show a corelation between memory/concentration/orientation disorders and Tramadol for both men and women. Thus, together they create a rule that is true for both genders. Rules 2.7 and 2.12 together create a correlation between relationship problem with partner and prescription of Tramadol. Rule 2.4 and 2.16 associate the problem with the partner's illness with the Tramadol prescription. Rules 2.10 and 2.15 show that both genders have a corelation between insomnia and Tramadol. Rules 2.1 and 2.2 create an association between dementia and the prescription of Tramadol, regardless of gender.

Table 7: Results from the TURS algorithm, including rules 'If ... Then', Probability scores and Coverage. First digit in `Rule ID` cells describe the dataset it comes from. Dataset 1 (ID 1.*) describes the dataset with ICPC code as the target label. Dataset 2 (ID 2.*) describes the dataset with ATC code as the target label.

Rule ID	**Features (if)**	**Target (then)**	**Probability**	**Coverage**
1.1	M=0; N02AB03=1	P06	0.1866	8736
1.2	V=0; N02AB03=1	P06	0.2101	3998
1.3	N02AE01=1	P06	0.1928	3750
1.4	M=0; N02AA01=1	P06	0.1682	7080
1.5	V=0; N02AA01=1	P06	0.1970	4117
1.6	M=1; N02AA05=1	P06	0.2026	12368
1.7	V=0; N02AX02=1	P06	0.1843	21885
1.8	M=1; N02AA05=0; N02AA59=0; N02AJ06=0; N02AJ13=0; N02AX02=0;	P06	0.1900	5583
1.9	M=1; N02AA05=0; N02AA59=0; N02AJ06=0; N02AX02=0; N02AX52=1	P06	0.1859	4636
1.10	V=0; N02AJ06=1	P06	0.1776	9869
1.11	V=0; N02AA59=1; N02AJ06=0; N02AX02=0	P06	0.1864	3546
1.12	N02AX02=1	P06	0.1664	65904
2.1	M=0; P70=1	N02AX02	0.2013	5247
2.2	M=1; P70=1	N02AX02	0.2279	2238
2.3	V= 1; P20=1	N02AX02	0.2453	8101
2.4	M=1; Z14=1	N02AX02	0.3026	2095
2.5	M=1;P20=1	N02AX02	0.2667	4589
2.6	M=0;Z15=1	N02AX02	0.2574	8026
2.7	M=1;Z12=1	N02AX02	0.3809	3211
2.8	M=1;Z15=1	N02AX02	0.2795	2683
2.9	M=1 ;P78=1	N02AX02	0.3713	2852
2.10	M=1;P06=1	N02AX02	0.3165	12911
2.11	M=0; P78=1	N02AX02	0.3486	7552
2.12	M=0;Z12=1	N02AX02	0.3455	6817
2.13	M=1;P17=1	N02AX02	0.3559	8701
2.14	M=0; P02=1	N02AX02	0.3252	10079
2.15	V=1;P06=1	N02AX02	0.2892	23879
2.16	M=0; Z14=1	N02AX02	0.2846	4501
2.17	P29=1	N02AX02	0.3061	6834
2.18	M=1; P01=1	N02AX02	0.3499	4881
2.19	P03=1	N02AX02	0.3129	11877
2.20	P02=0; P17=0; P76=1	N02AX02	0.3200	17785

Table 8: TURS Model metrics. Dataset 1 (ID 1) describes the dataset with ICPC code as the target label. Dataset 2 (ID 2) describes the dataset with ATC code as the target label.

ID	ROCAUC Test	ROCAUC Train	Gen Gap	LogLoss Test	LogLoss Train	Avg rule length	train test prob diff	overlap perc
1	0.535660	0.537767	0.002107	2.67237	2.67167	2.818	0.00198	0
2	0.544074	0.547459	0.003385	1.93261	1.93113	1.85	0.00456	0.000019

5.3 Comparison

Table 9 summarizes the results of the comparison. Out of 32 TURS rules, none of them were classified as an `Exact` match with the baseline rules. 7 rules (2.1, 2.2, 2.3, 2.5, 2.7, 2.12, 2.19) were identified as a `Close` match, which counts for approximately 21,9 % of the ruleset. 11 rules (1.7, 1.10, 2.6, 2.8, 2.9, 2.10, 2.13, 2.15, 2.18) were aligned on a `Partial` level, and 5 on a `Novel` level (1.1, 1.2, 1.6, 1.8, 1.9). 9 TURS rules were not covered by any rule of the Apriori and FP-growth set, which represents approximately 28,1 % of the instances.
Rules which are are unique to this method are as follows:

1.3 If Buprenorphine, then Insomnia/other sleep disorder (probability: 0.1928, coverage: 3750)

1.4 If not a man and Morphine, then Insomnia/other sleep disorder (probability: 0.1682, coverage: 7080)

1.5 If not a woman and Morphine, then Insomnia/other sleep disorder (probability: 0.1970, coverage: 4117)

1.11 If not a woman and Codeine, combinations excluding psycholeptics, then Insomnia/other sleep disorder (probability: 0.1864, coverage: 3546)

2.4 If man and Problem with partner's illness, then Tramadol (probability: 0.3026, coverage: 2095)

2.14 If not a man and Crisis/transient stress response, then Tramadol (probability: 0.3253, coverage: 10079)

2.16 If not a man and Problem with partner's illness, then Tramadol (probability: 0.2846, coverage: 4501)

2.17 If Other psychological symptoms/complaints, then Tramadol (probability: 0.3061, coverage: 6834)

2.20 If not Crisis/transient stress response; not Tobacco abuse; Depression, then Tramadol (probability: 0.3200, coverage: 17785)

The TURS rules that are closely aligned with the baseline rule(s) are as follows:

2.1 If not a man and Dementia, then Tramadol (probability: 0.2013, coverage: 5247)

2.2 If man and Dementia, then Tramadol (probability: 0.2279, coverage: 2238)

2.3 If woman and Memory/concentration/orientation disorders, then Tramadol (probability: 0.2453, coverage: 8101)

2.5 If man and Memory/concentration/orientation disorders, then Tramadol (probability: 0.2667, coverage: 4589)

2.7 If man and Relationship problem with partner, then Tramadol (probability: 0.3809, coverage: 3211)

2.12 If not a man and Relationship problem with partner, then Tramadol (probability: 0.3455, coverage: 6817)

2.19 If Feeling down/depressed, then Tramadol (probability: 0.3129, coverage: 11877)

Rules 2.3 and 2.4 together create an association between the Memory/concentration/orientation disorders and the Tramadol prescription for both genders, which then creates an `Exact` match with the baseline rule I103.

Table 9: Comparison of TURS rules with Apriori & FP-Growth rules. Rules that are the most aligned across the methods are marked with green color in the last column. TURS rules that did not match any rules from the baseline approach are marked with the `No matching baseline rule` message.

Begin of Table

TURS rule ID	**English description of rule**	**Matched baseline rule ID(s)**	**Align level**
1.1	If **man=0** and **Fentanyl=1**, then **Insomnia/other sleep disorder**.	I12	Novel
1.2	If **woman=0** and **Fentanyl=1**, then **Insomnia/other sleep disorder**.	I12	Novel
1.3	If **Buprenorphine=1**, then **Insomnia/other sleep disorder**.	No matching baseline rule	-
1.4	If **man=0** and **Morphine=1**, then **Insomnia/other sleep disorder**.	No matching baseline rule	-
1.5	If **woman=0** and **Morphine=1**, then **Insomnia/other sleep disorder**.	No matching baseline rule	-
1.6	If **man=1** and **Oxycodone=1**, then **Insomnia/other sleep disorder**.	I12	Novel
1.7	If **woman=0** and **Tramadol=1**, then **Insomnia/other sleep disorder**.	I22, I12	Partial

Continuation of Table

TURS rule ID	English description of rule	Matched baseline rule ID(s)	Align level
1.8	If **man=1** and **Oxycodone=0; Codeine, combinations excluding psycholeptics=0; Codeine and Paracetamol=0; Tramadol and Paracetamol=0; Tramadol=0;**, then **Insomnia/other sleep disorder**.	I12	Novel
1.9	If **man=1** and **Oxycodone=0; Codeine, combinations excluding psycholeptics=0; Codeine and Paracetamol=0; Tramadol=0; Tramadol combinations=1**, then **Insomnia/other sleep disorder**.	I13	Novel
1.10	If **woman=0** and **Codeine and Paracetamol=1**, then **Insomnia/other sleep disorder**.	I22	Partial
1.11	If **woman=0** and **Codeine, combinations excluding psycholeptics=1**, then **Insomnia/other sleep disorder**.	No matching baseline rule	-
1.12	If **Tramadol=1**, then **Insomnia/other sleep disorder**.	I12, I22	Partial
2.1	If **man=0** and **Dementia=1**, then **Tramadol**.	I121	Close
2.2	If **man=1** and **Dementia=1**, then **Tramadol**.	I121	Close
2.3	If **woman=1** and **Memory/concentration/orientation disorders=1**, then **Tramadol**.	I40, I103	Close
2.4	If **man=1** and **Problem with partner's illness=1**, then **Tramadol**.	No matching baseline rule	-
2.5	If **man=1** and **Memory/concentration/orientation disorders=1**, then **Tramadol**.	I40, I103	Close
2.6	If **man=0** and **Loss/death of partner=1**, then **Tramadol**.	I44	Partial
2.7	If **man=1** and **Relationship problem with partner=1**, then **Tramadol**.	I42, I75	Close
2.8	If **man=1** and **Loss/death of partner=1**, then **Tramadol**.	I44	Partial
2.9	If **man=1** and **Overvoltage=1**, then **Tramadol**.	I49, I107	Partial

Continuation of Table

TURS rule ID	English description of rule	Matched baseline rule ID(s)	Align level
2.10	If **man=1** and **Insomnia/other sleep disorder=1**, then **Tramadol**.	I33, I49	Partial
2.11	If **man=0** and **Overvoltage=1**, then **Tramadol**.	I49, I107	Partial
2.12	If **man=0** and **Relationship problem with partner=1**, then **Tramadol**.	I42, I75	Close
2.13	If **man=1** and **Tobacco abuse=1**, then **Tramadol**.	I32, I33, I41	Partial
2.14	If **man=0** and **Crisis/transient stress response=1**, then **Tramadol**.	No matching baseline rule	-
2.15	If **woman=1** and **Insomnia/other sleep disorder=1**, then **Tramadol**.	I42, I49	Partial
2.16	If **man=0** and **Problem with partner's illness=1**, then **Tramadol**.	No matching baseline rule	-
2.17	If **Other psychological symptoms/complaints=1**, then **Tramadol**.	No matching baseline rule	-
2.18	If **man=1** and **Feeling anxious/nervous/tense=1**, then **Tramadol**.	I41	Partial
2.19	If **Feeling down/depressed=1**, then **Tramadol**.	I68	Close
2.20	If **Crisis/transient stress response=0; Tobacco abuse=0; Depression=1**, then **Tramadol**.	No matching baseline rule	-

End of Table

6 Discussion

Dataset 1 contained a larger number of distinct target labels, as number of ICPC codes was higher. This resulted in the lower probabilities assigned to each individual class, compared to those observed in Dataset 2, where the number of ATC labels was smaller. `ROCAUC` scores for both datasets imply that the model generalizes better than a random guessing. However, they are not much better than that. This is likely due to the fact that there exist many possible target labels. While rules cover objectively large portions of the dataset, they are not strongly predictive. `Generalization Gap` scores in both cases are very low, which suggests that the model does not have the problem with overfitting, as performance is almost identical in the train and test data. `Average rule length` is better for the second dataset, however both models perform very well. Rules are short and thus easy to interpret. On the other hand, they might be too general, which correlates to low ROCAUC values. `Train - test probability difference` scores are close to zero, which means that predictions remained stable between the two stages. `Overlap` percentage scores are zero and almost zero, respectively. This makes the rules mutually exclusive and implies that each instance in

the dataset was covered by at most one rule. In conclusion, it can be understood that TURS ruleset contains short rules that do not overlap and do not overfit. However, ROCAUC score suggests that they do not capture strong predictive patterns.

Most rules assign the most frequent code in the dataset as their predicted label. This shows the influence of class imbalance on the TURS model, which aims to maximize coverage. In this case, it reduced the algorithm's ability to detect meaningful patterns for less frequent labels, which highlights the drawback of the TURS method. Another reason for the worse performance can be the size of the dataset. The initial research with TURS tested the algorithm on considerably smaller datasets (L. Yang & van Leeuwen, 2024). There exists a study on a larger dataset, yet it proved TURS to not be reliable. The main reason for this was the large size of the data and the skewed distribution of the target features (Peeters, 2025). It correlates to the characteristics of this study.

TURS output had significantly fewer rules in comparison to Apriori and FP Growth. This is due to the Minimum Description Length principle, which prioritizes compression. Unlike Apriori, which finds all frequent co-occurrences above a threshold, TURS selectively identifies only those rules that contribute to compressing the dataset. While Apriori and FP-growth returned over 130 rules, many of these were variations of the same underlying patterns, which did not happen with TURS implementation. An important feature of the TURS results is the presence of rules that include absence conditions, such as "if P06 = 0". Such patterns could not be generated by Apriori or FP-Growth, since those methods operate only on the item presence and do not encode the absence of attributes. This nuance contributed to the presence of new rules.

Data preprocessing played an important role in this study. While several methods were considered, the required input format of the TURS model made it a challenging. Several associations were found during the research. Prescription opioids such as Buprenorphie and Tramadol are often correlated with insomnia and other sleep disorders. This is an insightful association, as it has been identified in the previous baseline research on the same data. Moreover, there exist studies regarding this phenomenon (Cheatle & Webster, 2015) (Serdarevic, Osborne, Striley, & Cottler, 2017). A correlation was found between being a male, being prescribed Oxycodone, and struggling with insomnia. This rule scored coverage of 12368 and 0.2026 probability, which is a high score in the given dataset. Many associations have been found between the prescription of Tramadol and various psychological and social diagnosis. They include relationship problems with a partner, loss of a partner, depression, and memory disorders. There exist studies that connect the prescription of Tramadol to various problems with memory (Bassiony et al., 2017). This corelation was also found during the baseline ARM study. Moreover, there exist medical cases in which Tramadol has been used in treating depression caused by social loss (Rougemont-Bücking, Gamma, & Panksepp, 2017). Another association that was found connected being a man, tobacco abuse and Tramadol prescription. It scored a relatively high probability of 0.3559.

During the alignment comparison of the TURS, Apriori, and FP-Growth results, it became clear that while some rules were overlapping, TURS produced 9 new rules. Firstly, it identified a correlation between the problem with partner's illness and the Tramadol prescription. Secondly, there was an association between codeine and buprenorphine with sleep disorders. Thirdly, an association between the stress response and other psychological complaints with the Tramadol prescription was found.

7 Conclusion

This study investigated the application of a MDL-based approach TURS to the analysis of social and psychological aspects of opioid prescription data, and contrasted it with baseline association rule mining.

RQ: To what extent does an MDL-based rule-mining approach improve over baseline association rule mining in the analysis of social and psychological aspects of opioid prescription data?

MDL-based approach improves over the baseline association rule mining by producing a compact set of probabilistic rules that generalized above the random guessing threshold in the training and testing data. TURS enhanced interpretability and reduced redundancy which was present in Apriori and FP-growth. Direct comparison of performance was not possible between those methods, as TURS does not operate on the same performance metrics. However, an alternative comparison was performed by the rule alignment check, which concluded that 9 new rules were discovered by the TURS algorithm. The remaining 23 rules overlapped with the results from the baseline approach.

SQ1: How do the results from MDL method compare to those obtained during ARM in terms of understanding the social and psychological effects of prescription opioids?

ARM approach obtained over 130 results. Only a subset of those rules was deemed interesting and interpreted. MDL method provided a smaller set of rules, which lacked some of the insights gained through the ARM approach. TURS found correlations between Tramadol and various social and psychological conditions, as well as connected numerous opioids to the diagnosis of insomnia. ARM methods found some additional patterns, such as association of morphine or fentanyl with dementia, or assigning the problem of chronic alcohol abuse to men. On the other hand, TURS found new correlations between Tramadol and problems with partner's illness and stress response, or association between codeine and buprenorphine with sleep disorders.

SQ2: What are the advantages and limitations of MDL-based rule mining in identifying patterns within the ELAN dataset?

The primary advantage of the MDL approach is the generation of small, non-overlapping set of interpretable rules, which can make it easier for healthcare domain experts to validate them in the clinical context. Additionally, the performance metrics show that TURS performed very well in terms of average rule length, Generalization Gap or Test-Train probability difference. This means that it is a reliable model for probabilistic rule mining. Limitations of the MDL-based approach include sensitivity to the class imbalance, which in turn can cause the model to omit important, but rare occurrences in the data. TURS does not handle large datasets well, which also could be the reason for lower performance on the ROCAUC metric. Moreover, the required data input format for the TURS model did not suit the ELAN framework well. Therefore, while promising, TURS cannot be applied yet to large scale healthcare settings.

8 Further Research

There exist several ways in which this research can be extended. Future work should evaluate TURS on alternative healthcare datasets with a lower count of null values and more balanced class distributions. It would also be beneficial if there exist more variables between files that connected all information. This could enable the model to find a richer set of patterns while still keeping it

compact.
Opioid related research would be more insightful if it included richer patient details, such as country of origin, martial status, education level or profession. Currently, this inclusion was not possible due to the large amount of null values. Including them in the next research would be beneficial for a deeper understanding of this topic.
Furthermore, the study revealed that TURS is sensitive to its input. Due to the csv-only input format and the nature of ELAN dataset, it has been difficult to optimally prepare the data. Developing other input methods could potentially improve model's performance in finding the best rulesets.
Future work could also include a clinical evaluation performed by a GP. Their interpretation would be a valuable addition to the current work.
Lastly, this thesis focused only on the TURS implementation. It might be a good idea to compare this model's performance on medical datasets against other advanced ML algorithms, such as the MDL-based KRIMP. Such a comparison could provide valuable insights on MDL-based models and the role of advanced machine learning in the healthcare field.

References

Al-Hasani, R., & Bruchas, M. R. (2011). Molecular mechanisms of opioid receptor-dependent signaling and behavior. *Anesthesiology*, *115*(6), 1363.

Aljaaf, A. J., Al-Jumeily, D., Hussain, A. J., Dawson, T., Fergus, P., & Al-Jumaily, M. (2015). Predicting the likelihood of heart failure with a multi level risk assessment using decision tree. In *2015 third international conference on technological advances in electrical, electronics and computer engineering (taeece)* (pp. 101–106).

Asadi, B., & Varadharajan, V. (2019). An mdl-based classifier for transactional datasets with application in malware detection. *arXiv preprint arXiv:1910.03751*.

Baker, A. (2022). Simplicity.'stanford encyclopedia of philosophy.

Bassiony, M. M., Youssef, U. M., Hassan, M. S., El-Deen, G. M. S., El-Gohari, H., Abdelghani, M., ... Ibrahim, D. H. (2017). Cognitive impairment and tramadol dependence. *Journal of clinical psychopharmacology*, *37*(1), 61–66.

Bentsen, B. G. (1986). International classification of primary care. *Scandinavian journal of primary health care*, *4*(1), 43–50.

Cheatle, M. D., & Webster, L. R. (2015). Opioid therapy and sleep disorders: risks and mitigation strategies. *Pain Medicine*, *16*(suppl_1), S22–S26.

Chenaf, C., Kaboré, J.-L., Delorme, J., Pereira, B., Mulliez, A., Zenut, M., ... Authier, N. (2019). Prescription opioid analgesic use in france: trends and impact on morbidity–mortality. *European Journal of Pain*, *23*(1), 124–134.

Dhalla, I. A., Persaud, N., & Juurlink, D. N. (2011). Facing up to the prescription opioid crisis. *Bmj*, *343*.

Dowell, D., Haegerich, T. M., & Chou, R. (2016). Cdc guideline for prescribing opioids for chronic pain—united states, 2016. *Jama*, *315*(15), 1624–1645.

Dydyk, A. M., Jain, N. K., & Gupta, M. (2024). Opioid use disorder: Evaluation and management. In *Statpearls*. Treasure Island, FL: StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK553166/. ([Updated 2024 Jan 17; Cited 2025 Jan])

Ellis, R. J., Wang, Z., Genes, N., & Ma'ayan, A. (2019). Predicting opioid dependence from electronic health records with machine learning. *BioData mining*, *12*(1), 3.

for Disease Control, C., Prevention, et al. (2013). *Unintentional drug poisoning in the united states. july 2010.*

Galbrun, E. (2022). The minimum description length principle for pattern mining: A survey. *Data mining and knowledge discovery*, *36*(5), 1679–1727.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... others (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *jama*, *316*(22), 2402–2410.

Han, D.-H., Lee, S., & Seo, D.-C. (2020). Using machine learning to predict opioid misuse among us adolescents. *Preventive medicine*, *130*, 105886.

Kalkman, G. A., Kramers, C., van Dongen, R. T., van den Brink, W., & Schellekens, A. (2019). Trends in use and misuse of opioids in the netherlands: a retrospective, multi-source database study. *The Lancet Public Health*, *4*(10), e498–e505.

Lo-Ciganic, W.-H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K., ... others (2019). Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA network open*, *2*(3), e190968–e190968.

Lu, S., Christie, G. A., Nguyen, T. T., Freeman, J. D., & Hsu, E. B. (2022). Applications of artificial intelligence and machine learning in disasters and public health emergencies. *Disaster medicine and public health preparedness*, *16*(4), 1674–1681.

Network, D. A. W. (2012). *Trends in emergency department visits involving non-medical use of narcotic pain relievers.*

Peeters, L. (2025). *Truly unordered rule sets for interpretable risk estimation in the retail sector* (Master's thesis). Leiden University, Leiden Institute of Advanced Computer Science (LIACS).

Pergolizzi Jr, J. V., Raffa, R. B., & Rosenblatt, M. H. (2020). Opioid withdrawal symptoms, a consequence of chronic opioid use and opioid use disorder: Current understanding and approaches to management. *Journal of clinical pharmacy and therapeutics*, *45*(5), 892–903.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*(5), 465–471.

Rougemont-Bücking, A., Gamma, F., & Panksepp, J. (2017). Use of tramadol in psychiatric care: a comprehensive review and report of two cases. *Swiss medical weekly*, *147*(1920), w14428–w14428.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, *1*(5), 206–215.

Scherrer, J. F., Salas, J., Copeland, L. A., Stock, E. M., Ahmedani, B. K., Sullivan, M. D., ... Lustman, P. J. (2016). Prescription opioid duration, dose, and increased risk of depression in 3 large patient populations. *The Annals of Family Medicine*, *14*(1), 54–62.

Schubert, I., Ihle, P., & Sabatowski, R. (2013). Increase in opiate prescription in germany between 2000 and 2010: a study based on insurance data. *Deutsches Ärzteblatt International*, *110*(4), 45.

Segal, Z., Radinsky, K., Elad, G., Marom, G., Beladev, M., Lewis, M., ... Koren, G. (2020). Development of a machine learning algorithm for early detection of opioid use disorder. *Pharmacology Research & Perspectives*, *8*(6), e00669.

Serdarevic, M., Osborne, V., Striley, C. W., & Cottler, L. B. (2017). The association between insomnia and prescription opioid use: results from a community sample in northeast florida. *Sleep Health*, *3*(5), 368–372.

Set, T. E. D. (n.d.). Substance abuse treatment admissions by primary substance of abuse, according to sex, age group, race, and ethnicity, funded by the substance abuse and mental health services administration, dhhs. 2003. *The latest data are available at*, 800–729.

Spruit, M. (2022, April). *Translational data science in population health.* Retrieved from https://doi.org/10.5281/zenodo.7665858 doi: 10.5281/zenodo.7665858

Sun, M., Chen, W.-M., Wu, S.-Y., & Zhang, J. (2023). Long-term opioid use and dementia risk in patients with chronic pain. *Journal of the American Medical Directors Association*, *24*(9), 1420–1426.

Taylor, J. L., & Samet, J. H. (2022). Opioid use disorder. *Annals of Internal Medicine*, *175*(1), ITC1–ITC16.

Tumkur Rameshchandra, R. (2024). *Unsupervised machine learning methods to understand the social and psychological effects of prescription opioids* (Master's thesis). Leiden University, Leiden Institute of Advanced Computer Science (LIACS).

Vanetik, N., & Litvak, M. (2018). Drim: Mdl-based approach for fast diverse summarization. In *2018 ieee/wic/acm international conference on web intelligence (wi)* (pp. 660–663).

van Leeuwen, M., Vreeken, J., & Siebes, A. (2006). Compression picks item sets that matter. In *European conference on principles of data mining and knowledge discovery* (pp. 585–592).

Van Zee, A. (2009). The promotion and marketing of oxycontin: commercial triumph, public health tragedy. *American journal of public health*, *99*(2), 221–227.

Webster, L. R. (2017). Risk factors for opioid-use disorder and overdose. *Anesthesia & analgesia*, *125*(5), 1741–1748.

Webster, L. R., Choi, Y., Desai, H., Webster, L., & Grant, B. J. (2008). Sleep-disordered breathing and chronic opioid therapy. *Pain Medicine*, *9*(4), 425–432.

Wirth, R., & Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1–10).

Yang, G., Ren, Y., Pan, Q., Ning, G., Gong, S., Cai, G., ... Yan, J. (2010). A heart failure diagnosis model based on support vector machine. In *2010 3rd international conference on biomedical engineering and informatics* (Vol. 3, pp. 1105–1108).

Yang, L., & van Leeuwen, M. (2024). Probabilistic truly unordered rule sets. *arXiv preprint arXiv:2401.09918*.

Zin, C. S., Chen, L.-C., & Knaggs, R. D. (2014). Changes in trends and pattern of strong opioid prescribing in primary care. *European journal of pain*, *18*(9), 1343–1351.