Evaluating Frequent-Set Mining Approaches in Machine-Learning Problems with Several Attributes: A Case Study in Healthcare

Shruti Kaushik[1(✉)], Abhinav Choudhury[1(✉)], Nataraj Dasgupta[2],
Sayee Natarajan[2], Larry A. Pickett[2], and Varun Dutt[1(✉)]

[1] Applied Cognitive Science Laboratory, Indian Institute of Technology Mandi,
Mandi 175005, Himachal Pradesh, India
{shruti_kaushik,
abhinav_choudhury}@students.iitmandi.ac.in,
varun@iitmandi.ac.in
[2] RxDataScience, Inc., New York 27709, USA
{nd,sayee,larry}@rxdatascience.com

Abstract. Often datasets may involve thousands of attributes, and it is important to discover relevant features for machine-learning (ML) algorithms. Here, approaches that reduce or select features may become difficult to apply, and feature discovery may be made using frequent-set mining approaches. In this paper, we use the Apriori frequent-set mining approach to discover the most frequently occurring features from among thousands of features in datasets where patients consume pain medications. We use these frequently occurring features along with other demographic and clinical features in specific ML algorithms and compare algorithms' accuracies for classifying the type and frequency of consumption of pain medications. Results revealed that Apriori implementation for features discovery improved the performance of a large majority of ML algorithms and decision tree performed better among many ML algorithms. The main implication of our analyses is in helping the machine-learning community solves problems involving thousands of attributes.

Keywords: Apriori algorithm · Frequent-set mining · Machine learning
Pain medications · Features

1 Introduction

Since the early 90s, machine-learning (ML) algorithms have been used to help mine patterns in data sets concerning fraud detection and others [1]. In recent years, ML algorithms have also been utilized in the healthcare sector [2]. In fact, the existence of electronic health records (EHRs) has allowed researchers to apply ML algorithms to learn hidden patterns in data to improve patient outcomes like the type of medications patients consume and the frequency at which they consume these medications [3]. Mining hidden patterns in healthcare data sets could help healthcare providers and pharmaceutical companies to plan quality healthcare for patients in need.

To predict healthcare outcomes accurately, ML algorithms need to focus on discovering appropriate features from data [4]. In general, healthcare data sets are large, and they may contain several thousands of attributes to enable learning of patterns in data [5]. The presence of many attributes in data sets may make it difficult to discover the most relevant features for predicting outcomes via ML algorithms.

Presence of thousands of attributes in the data is problematic for classification algorithms for processing these attributes as features require large memory usage and high computational costs [37]. Two techniques have been suggested in the literature to address the problem of datasets possessing a large number of features: feature reduction (dimensionality reduction) and feature selection [38]. Feature reduction technique reduces the number of attributes by creating new combinations of attributes; whereas, feature selection techniques include and exclude attributes present in the data without changing them [38]. Popular algorithms like Principal Component Analysis (PCA; for features reduction) and Analysis of variance (ANOVA; for features selection) have been used for datasets with a large number of features in the past [6, 7]. PCA is a linear feature-based approach that uses eigenvector analysis to determine critical variables in a high dimensional data without much loss of information [8]. ANOVA is a collection of statistical models used to analyze the differences between group means and their associated procedures (such as "variation" among and between groups) [9]. In ANOVA, the features that describe the most substantial proportion of the variance are the features that are retained in data [10]. Although both PCA and ANOVA approach seem to help in feature discovery, these approaches may become computationally expensive to apply in problems where there are thousands of features in data (e.g., thousands of diagnostic and procedure codes across several patient cases in medical datasets). Another disadvantage of the PCA method is that it is an elimination technique that considers a single feature to be important or unimportant to the problem rather than a group of features being important [8]. Similarly, in ANOVA, researchers need to test assumptions of normality and independence, which may not be the case when features depend upon each other [9, 10]. One way to address the challenge posed by data sets with several thousands of features is by using frequent item-set mining algorithms (e.g., Apriori algorithm) to discover a subset of features because these algorithms look at the associations among items while selecting frequent item-sets [11]. The primary goal of this paper is to evaluate the potential of Apriori frequent item-set mining algorithm for feature discovery before application of different ML algorithms. Specifically, we take a healthcare dataset involving consumption of two pain medications in the US, and we apply different ML algorithms both with and without a prior feature-discovery process involving the Apriori algorithm. The Apriori algorithm works on the fundamental property that an item-set is frequent only if all its non-empty subsets are also frequent [11]. Using the Apriori algorithm, we generate frequently appearing diagnosis and procedure codes in a healthcare dataset. Then, using these frequently occurring diagnosis and procedure codes as present/absent features, along with other features, we apply certain supervised ML algorithms. We check the benefits of using the Apriori algorithm by comparing the classification accuracies of certain ML algorithms when all attributes are considered as features in the dataset and when only the discovered attributes via Apriori are considered as features. To get confidence in our results, we replicate our analyses using several ML algorithms such as the decision

tree [27], Naïve Bayes classifier [30, 41], logistic regression [31], and support vector machine [33, 42].

In what follows, we first provide a brief review of related literature. In Sect. 3, we explain the methodology of using the Apriori algorithm for features discovery. In Sect. 4, we present our experimental results and compare classification accuracies for cases with or without the Apriori implementation. We conclude our paper and provide a brief discussion on the implication of this research and its future scope.

2   Background

Data analytics could help in providing useful insights in very large healthcare datasets [12]. These insights may aid in effective decision-making and save lives [12]. Researchers have applied many such techniques to mine the hidden knowledge in the medical domain [15]. For example, various data-mining approaches such as classification, clustering, statistical approaches, and association-rule mining have made a significant contribution to healthcare research [13, 15].

Healthcare researchers have used different ML algorithms to investigate research questions in healthcare [20–22]. For example, some researchers have used the Naive Bayes classification algorithm to diagnose heart diseases [20]. Others have used ML techniques like J48, MLP, Random Forest, SVM, and Bayesian Network classifiers to classify liver-disease patients [21]. Researchers have also used frequent-set mining approaches in the recent past [14–18]. For example, in the literature on frequent-set mining, some researchers have introduced the frequent-set mining pincer-search algorithm to discover the maximum frequent-item sets [15, 17]. Similarly, Rani, Prakash, and Govardhan [16] presented a model for multilevel association rule mining, which satisfies the different minimum support at each level. Also, certain researchers have focused on identifying frequent diseases using the frequent-set mining algorithms like Apriori [14] and mined different association rules for consumers of certain pain medications [19].

Furthermore, prior research has combined methods like PCA with ML algorithms for feature extraction and subsequent classification. For example, certain researchers have used traditional methods like PCA, rough PCA, unsupervised quick reduction algorithm, and empirical distribution ranking approach to extract features that could be further used for an ML classification task [22]. However, to the best of authors' knowledge, prior research has yet to combine frequent-set mining algorithms along with ML algorithms in feature-discovery and predicting healthcare outcomes. In this paper, we attend to this literature gap and apply the Apriori frequent-set mining algorithm for features discovery before performing machine learning for classifying healthcare outcomes. To test our Apriori approach, we take certain healthcare dataset where there are potentially thousands of features to choose between and where feature selection via traditional methods may become computationally expensive. We investigate the performance of different ML algorithms with and without frequent-set mining using the Apriori approach.

3 Method

3.1 Data

In this paper, we use the Truven MarketScan® health dataset containing patients' insurance claims in the US [23]. The data set contains 120,000 patients, who consumed two pain medications, medicine A, medicine B, or both between January 2011 and December 2015.[1] The dataset contains patients' demographic variables (age, gender, region, and birth year), clinical variables (admission type, diagnoses made, and procedures performed), the name of medicines, and medicines' refill counts per patient. We used a big-data architecture consisting of q-programming language to query a kdb+ database [24] for fetching patient records who have consumed pain medication A, B, or both during their journeys. After fetching data from the kdb+ database, we prepared a file containing different diagnoses and procedures corresponding to different patients, who consumed both medications. The dataset contains 55.20% records of patients who consumed medicine A only, 39.98% records of medicine B only, and 4.82% records for those patients who consumed both these medications. There were 15,081 attributes present in total against each patient in this dataset. These attributes consist of patients' age, gender, region, type of admission, diagnoses and procedures performed on the patient, medicine name and its refill information. Out of 15,081 attributes, 15,075 attributes were diagnoses and procedure codes some of which were inter-related.

The diagnoses and procedures were written for patients using the International Classification of Diseases (ICD)-9 codes [25]. The ICD codes are used by physicians and other healthcare providers to classify different diagnoses and procedures recorded during different illnesses in the United States [25]. We applied the Apriori frequent-set mining algorithm to diagnoses made and procedures performed for different patients consuming the two pain medications. The Apriori algorithm discovered the frequently appearing diagnoses and procedures among the 15,075 unique diagnoses procedure codes available in the dataset. We used these frequently occurring diagnoses and procedures as input features along with other independent variables in different ML algorithms that were applied to our dataset. The ML algorithms classified patients according to the type of medications consumed and the frequency of refilling different medications in the dataset.

3.2 Association-Rule Mining

Association-rule mining [36] is a popular technique that aims to extract associations among items in data. An association rule represents a relationship between a group of objects in the database. The basic model of association-rule mining is described below.

Let I = {I_1, I_2, …, I_n} be a set of n distinct items, where each attribute I_1, I_2 … I_n is binary (0 or 1) in nature. T is a transaction with a unique transaction id that contains a subset of items in I. Let D be a database having different transaction records, where each transaction is differentiated by the transaction ID and may contain a subset of

---

[1] Due to a non-disclosure agreement, we have anonymized the actual names of these medications.

items in I. Thus, D = {T_1, T_2, …, T_M}. An association rule is an implication in the form of X → Y, where X, Y ⊆ I and X ∩ Y = ∅. X is called an antecedent while Y is called consequent. There are two measures for finding association rules: support (S) and confidence (C). Support of an item-set X is defined as the proportion of the transactions that contain the item-set X in the database D.

$$Support(X) = \frac{count(X)}{count(D)} \tag{1}$$

The confidence of an association rule X → Y is defined as the proportion of transactions that contain both items X and Y in all the transactions that contain item X.

$$Confidence(X \rightarrow Y) = P(Y \mid X) = \frac{Support(X \cup Y)}{Support(X)} \tag{2}$$

The confidence is a measure of the strength of the association rules. If there is an association rule "X → Y" whose support and confidence satisfies minimum support threshold (min_support) and minimum confidence threshold (min_conf) provided by a user, then we call it an association rule with min_support and min_conf.

3.2.1 Apriori Algorithm

The Apriori algorithm [11] finds frequent item-sets using an iterative level-wise approach based on candidate generation. This algorithm works in following steps:

1. The transactions in database D are scanned to determine frequent 1-itemsets, L_1 that possess the minimum support.
2. Generate candidate k item-sets C_k from joining two k − 1 itemsets, L_{k-1}, and remove its infrequent subset.
3. Scan D to get support count for each k item-sets, C_k.
4. The set of frequent k item-sets, L_k, is then determined. L_k results from support count of candidate k − 1 item-sets.
5. Back to step 2 until there is no candidate k + 1 item-sets, C_{k+1}.
6. Extract the frequent k item-sets, L = L_k.

The Apriori algorithm was used to mine the frequent diagnoses and procedures out of 15,075 unique diagnoses and procedures present in the dataset. After getting the 9 frequent diagnoses and procedures from the result of Apriori algorithm, we formed two ML problems. In the first problem, we used the frequent diagnoses and procedure codes as categorical variables along with the other 6-independent variables to classify patients by the medication they consumed; i.e., consuming medicine A, B, or both. Thus, the first ML problem is a three-class problem. In the second ML problem, we used the frequent diagnoses and procedure codes obtained from the result of Apriori algorithm along with the other 6-independent variables to classify patients as frequent or infrequent buyers of medications they consumed. This ML problem is a two-class classification problem.

We then applied different ML algorithms to all 15,081 attributes in the dataset and compared the classification results with the case when only 15 attributes were used in the ML algorithms. Next, we discuss the brief descriptions of different ML algorithms that we used in this paper, i.e., decision tree [27], Naïve Bayes [30, 41], logistic regression [31], and support vector machine [33, 42]. We applied ZeroR [26] as a base line algorithm to compare the classification accuracies of different ML algorithms mentioned above.

3.3   Machine-Learning Algorithms

3.3.1   ZeroR

ZeroR is the simplest classification method which relies on the target to be classified and ignores all predictors [26]. The ZeroR classifier simply predicts all points as belonging to the majority class. Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a benchmark for other classification methods.

3.3.2   Decision Tree

A decision tree is a tree, where non-leaf nodes denote tests on attributes, each branch denotes the result of different tests, and each leaf node denotes the class label [27]. In the decision tree, each internal node is labelled with an input feature, and leaf of the tree either gives a class label or a probability distribution over the classes. The results obtained from decision trees are easier to interpret. Following assumptions are taken into account while creating a decision tree [28]:

1. Initially, the complete training set is considered as the root.
2. Feature values are preferred to be categorical. If the values are continuous, then they are discretized before building the model.
3. Records are distributed recursively by attribute values.
4. Order of placing attributes as root or internal node of the decision tree is done by using a statistical approach based on the calculation of entropy and gain.

The first challenge in a decision tree implementation is to identify which of the attributes to select as the root node and at each level. Random selection of nodes may give bad results with very low accuracy. Handling this problem is known as the attributes selection. We have used the information-gain measure to identify the attribute that can be considered as the root node at each level [39].

3.3.3   Naïve Bayes

This classifier belongs to a family of probabilistic classifiers that is based on the Bayes theorem with strong independence assumptions between features [30, 41]. It assumes that the value of a particular feature is independent of the value of any other feature, given the class variable. This classifier attempts to maximize the posterior probability in determining the class of a transaction. A Naïve Bayes classifier assumes features to contribute independently to the probability, regardless of any correlation between features.

Suppose, n be the number of features in a problem which is represented by a vector $y = (y_1, y_2, \ldots, y_n)$ and K be the possible number of classes C_k. Naïve Bayes is a conditional probability model which can be decomposed as [40]:

$$p(C_k/y) = \frac{p(C_k)p(y/C_k)}{p(y)} \qquad (3)$$

In practice, the numerator of this equation determines the LHS (the denominator is a constant). Under the independence assumption among attributes, the probability of certain attributes belonging to a certain class is defined as [40]

$$p(C_k/y_1, \ldots, y_n) = p(C_k) \prod_{i=1}^{n} p(y_i/C_k) \qquad (4)$$

A common rule is to pick the class that is the most probable. This most probable class is defined by the maximum a posteriori (MAP) decision rule [40] as:

$$y = argmax_{k \in 1 \ldots K} p(C_k) \prod_{i=1}^{n} p(y_i/C_k) \qquad (5)$$

3.3.4 Logistic Regression

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables [31]. The dependent variable in logistic regression or logit model is categorical. Logistic regression is named for the function used at the core of the method, the logistic function [32]. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits. Input values (x) are combined linearly using weights or coefficient values to predict an output value (y). The simple logistic regression is defined as:

$$y = e^{(b0 + b1*x)} / \left(1 + e^{(b0 + b1*x)}\right) \qquad (6)$$

where y is the predicted output, b0 is the bias or intercept term, and b1 is the coefficient for the single input value (x). In the general logistics regression model, each column in data has an associated b coefficient (a constant real value) that must be learned from the training data.

3.3.5 Support Vector Machines

Support vector machines are supervised learning models that are binary classification algorithms [33]. Support vector machines construct a hyperplane or a set of hyperplanes in a high dimensional space that can be used for classification, regression, or other tasks. There are two types of support vector machines: linear SVM and the non-linear SVM [42]. If the data is linearly separable, then the linear SVM is sufficient to perform classification. However, if the problem cannot be classified linearly, then we

require a non-linear SVM to perform the task. The non-linear support vector machine function takes the data into the high dimensional plane and then performs classification [42]. In the SVM algorithm, we optimize the support weights to minimize the objective (error) function for better classifications.

3.4   Model Calibration

3.4.1   Apriori Implementation

To find the frequent diagnoses and procedure codes using Apriori algorithm, we needed to set the threshold limits for support and confidence. For setting the minimum threshold limits for support and confidence, we conducted sensitivity analyses. We first calculated the male-to-female ratio among patients who consumed either medication A, B, or both in the dataset. This ratio turned out to be 0.58. Then, we tried different values of minimum support and minimum confidence till the point when the association rules obtained from Apriori algorithm had the same male-to-female ratio of 0.58 as in the dataset. With 3% threshold support and 99% threshold confidence, we got the similar male-to-female ratio for all the association rules with frequent diagnoses and procedure codes. We got three association rules when we applied the Apriori algorithm. These rules were made up of the following nine ICD-9 codes: total knee arthroplasty, osteoarthrosis secondary lower leg, removal of foreign body from the eye, total knee replacement, osteoarthrosis primary lower leg, osteoarthrosis generalized lower leg, total hip arthroplasty, iridectomy, and total hip replacement.

3.4.2   Machine Learning Combined with Apriori

For the ML analyses, the dataset was randomly divided into two parts: 70% of the data was used for training, and 30% of the data was used for testing. We used the d-prime = z (true-positive rate) – z (false-positive rate) as measure of accuracy [34, 43]. The higher the d-prime, the better the performance (a d-prime = 0 indicates random performance, where true-positive rate = false-positive rate). Our first machine learning problem is a three-class problem, where we classified a patient according to the medication consumption. So, a patient can be classified under class A, class B or both. Our second ML problem is a two-class problem, where we classified patients according to their frequency of medicine consumption. So, a patient can be classified as a frequent buyer or an infrequent buyer of medications. We took the median of refill counts per patient (=3) to distinguish between a frequent buyer (>3) and an infrequent buyer ($\leq$ 3). In the dataset, 41.11% patients belong to the frequent class and rest to the infrequent class. We used the frequent codes obtained from Apriori algorithm as categorical variables along with the other demographic and clinical features while training our ML models. For different classification problems, we used different features from the original dataset and the Apriori output (see Table 1). Table 1 shows the list of 15-features used in different ML models after applying Apriori procedure for the three-class and two-class classification problems. As shown in Table 1, some of these 15-features were excluded in certain problems (e.g., refill count was excluded in the

Table 1. Description of input features for classification problems and their source

Features	Description	Features for 3-class problem	Features for 2-class problem
Sex	Male Female	Included	Included
Age group	0–17, 18–34, 35–44, 45–54, 55–64	Included	Included
Region	Northeast, northcentral, south, west, unknown	Included	Included
Type of admission	Surgical, medical, maternity and newborn, psych and substance abuse, unknown	Included	Included
Refill count	Count in number	Included	Excluded
Pain medication	A, B, Both	Excluded	Included
Total knee arthroplasty	Present/not present	Included	Included
Osteoarthrosis of secondary lower leg	Present/not present	Included	Included
Removal of foreign body from eye	Present/not present	Included	Included
Total knee replacement	Present/not present	Included	Included
Osteoarthrosis of primary lower leg	Present/not present	Included	Included
Osteoarthrosis of generalized lower leg	Present/not present	Included	Included
Total hip arthroplasty	Present/not present	Included	Included
Iridectomy	Present/not present	Included	Included
Total hip replacement	Present/not present	Included	Included

two-class problem). That is because these attributes were strongly correlated with the predicted class. Certain features of this table were directly included from the dataset (sex, age group, region, type of admission, refill count and pain medication) while rest were included after applying Apriori on the procedure and diagnoses codes.

For both the classification problems, we ran two different datasets on different ML models. The first dataset contained features obtained from the Apriori algorithm along with other demographic and clinical features from the dataset; whereas, the second dataset contained all the features (=15,081).

4 Results

4.1 Apriori Algorithm

Based on the Apriori algorithm [11], we found the following three association rules among patients who consumed either medicine A, medicine B, or both during their patient journeys:

1. *If a patient goes for **total knee arthroplasty**, **osteoarthrosis of primary/secondary/generalized lower leg** and **removal of foreign body from posterior eye segments**, then he/she goes **for total knee replacement** and consumes A/B.*
2. *If a patient goes for total knee arthroplasty, then he/she goes for total knee replacement and consumes A/B.*
3. *If a patient goes for **total hip arthroplasty** and **Iridectomy**, then he/she goes for **total hip replacement** and consumes both the medications.*

As explained above, we took nine frequently occurring ICD-9 diagnoses and procedures codes from the rules above. These codes have been bolded in the rules.

4.2 Machine-Learning Algorithms

We applied various ML algorithms like Naïve Bayes, Decision Tree, Logistic Regression, Support Vector Machine (linear kernel), and Support Vector Machine (radial kernel) [26–33] on our dataset and compared their classification accuracy using d-prime. Figure 1 shows the d-prime results from different ML algorithms for the three-class problems (Fig. 1A) and two-class problems (Fig. 1B) with and without Apriori algorithm implementation. We have only used 1000 features for SVMs as the algorithm was not able to scale for 15,081 features (in without Apriori implementation). In the three-class problem (Fig. 1A), without the Apriori implementation, the best d-prime was obtained by the SVM with a radial kernel function. However, with the Apriori implementation in the three-class problem, the best performance was obtained by the decision tree. The performance of all the algorithms improved with the implementation of Apriori algorithm. Furthermore, in the two-class problem (Fig. 1B), both with and without the Apriori implementation, the best d-prime was obtained by the decision tree. In fact, barring the decision tree, all other algorithms possessed a d-prime = 0 (true-positive rate = false-positive rate) in the two-class problem without the Apriori implementation. Barring the SVM with linear kernel and ZeroR algorithms, the performance of all other algorithms improved with the implementation of Apriori algorithm. In general, for both with and without Apriori implementation, all algorithms performed better (higher d-prime) in the three-class problem compared to the two-class problem.
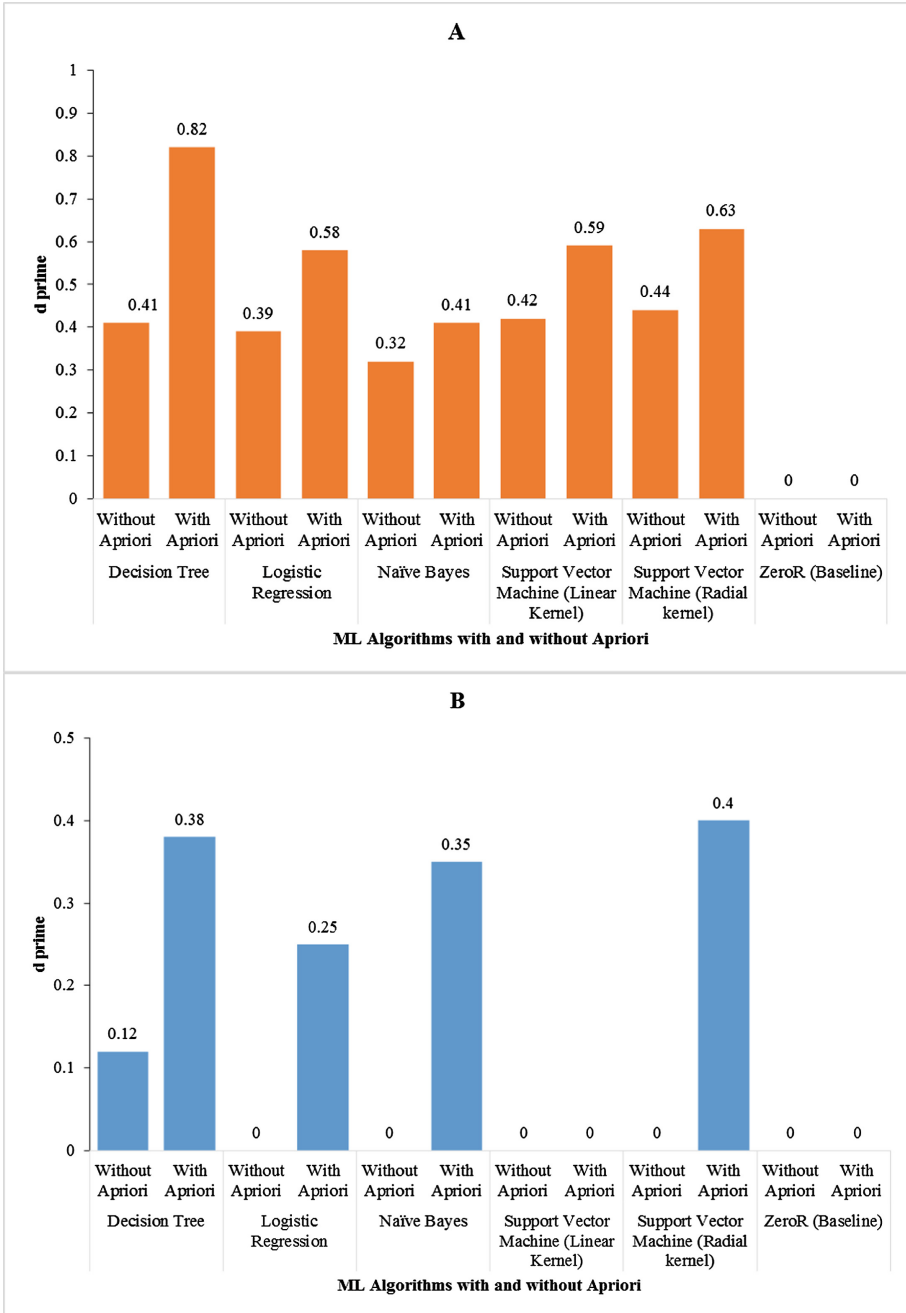
Fig. 1. The d-prime results from different ML algorithms for the three-class problems (A) and two-class problems (B) with and without Apriori algorithm implementation.

5   Discussion and Conclusions

Medical data concerning patient and their journeys may likely contain a large number of attributes detailing demographics as well as procedural and diagnostic information [44, 45]. Thus, before attempting different machine-learning (ML) algorithms on patient-journey datasets, it would be good to reduce the number of attributes used as features. One way to address this reduction is by using frequent item-set mining algorithms (e.g., Apriori algorithm). These algorithms help discover a subset of features by evaluating the associations among items in different transactions [11]. The primary objective of this paper was to evaluate the potential of Apriori frequent item-set mining algorithm for features discovery before application of different ML algorithms. First, using Apriori algorithm, we discovered the frequently occurring attributes (diagnoses and procedures) among patients consuming pain medications. The Apriori frequent-set mining approach gave nine frequently occurring diagnoses and procedure attributes in association rules out of a total of 15,075 possible attributes in the dataset. Second, we found that the Apriori implementation led to improved performance from ML algorithms: in general, the d-prime was higher after application of Apriori compared to when Apriori was not applied, and all attributes were considered in the algorithms. Third, we found that the ML algorithms classified the patients according to the medication used (the three-class problem) better compared to the frequency of medication used (the two-class problem). Finally, we found that the decision-tree algorithm performed better compared to a large number of ML algorithms across both the three-class and two-class problems.

First, we found that using the Apriori implementation [11] improved the performance of a large majority of ML algorithms. One likely reason for this finding is that Apriori procedures allow us to find features that frequently occur together or are correlated with each other. For example, the nine-attributes selected by the Apriori algorithm out of a total of 15,075 attributes occurred in three association rules that possessed the confidence of 99%. Given the high confidence of these rules, the attributes present in them were highly correlated. As the rules predicted the use of medications, these attributes seemed to predict the medications' use and their frequency well. Overall, given our results, the Apriori algorithm seems to be a suitable technique for identifying important features, when the dataset contains thousands of relevant or irrelevant attributes.

Third, we found that the ML algorithms classified the patients according to the medication used (the three-class problem) better compared to the frequency of medication used (the two-class problem). One likely reason for this finding could be the nature of the predicted class in the three-class problem compared to the two-class problem. In the three-class problem, the predicted class was the medicine name, which is a discrete attribute. However, in the two-class problem, the predicted class was the frequency of the medicine use, which is a discrete attribute derived from a continuous attribute (frequency/refill count). On account of the differences between the nature of the predicted classes, it seems that the classification boundary could divide one class from the other in the three-class problem compared to the two-class problem. However,

this line of reasoning needs to be further explored for other predicted classes (discrete or continuous) as well as for other datasets.

Overall, we found that the decision tree algorithm performed better compared to a large class of algorithms across both the 2- and 3- class classification problems. Decision tree implicitly performs feature selection using measures like information gain, and they do not require making any assumption regarding linearity in the data [35]. Thus, it seems that decision trees can classify datasets where there are a large number of relevant and irrelevant attributes [28]. Also, due to the superior performance of the decision tree algorithm, we believe that this algorithm could be used for performing machine-learning in healthcare datasets.

Our results have some important real-world implications for using data analytics in the healthcare domain. First, we believe that supervised learning algorithms like decision trees and others can classify medicine intake and the medicine frequency to a certain accuracy. However, as the d-prime values were all less than 1.0 in our results, we believe that these algorithms need more improvements before we can reliably use them for confirming hypotheses in healthcare datasets. Based upon our results, among different algorithms, we would suggest decision trees to be most robust in datasets with a large number of attributes. Second, we also believe that predicting the type of medications consumed and their frequency of use could be extremely helpful for pharmaceutical companies to decide upon their drug manufacture strategies. Specifically, such strategies could reduce supply-chain costs by managing the delays in ordering and stocking of medications.

In this paper, we performed a preliminary analysis on using frequent-set mining approach on a healthcare dataset involving several attributes, and there are a number of extensions possible of this work in the near future. For example, in future, we would like to compare the classification accuracies of ML algorithms after applying PCA and ANOVA and other features selection techniques to those resulting from applying frequent-set mining algorithms. Here, it would be interesting to extend our investigation to other medical and non-medical datasets as well as to other ML algorithms (e.g., neural networks). We plan to embark on some of these ideas as part of our research in the near future.

References

1. Seeja, K.R., Zareapoor, M.: FraudMiner: a novel credit card fraud detection model based on frequent itemset mining. Sci. World J. (2014)
2. Oswal, S., Shah, G., Student, P.G.: A study on data mining techniques on healthcare issues and its uses and application on health sector. Int. J. Eng. Sci. **7**, 13536 (2017)
3. Parikh, R.B., Obermeyer, Z., Bates, D.W.: Making Predictive Analytics a Routine Part of patient Care. https://hbr.org/2016/04/making-predictive-analytics-a-routine-part-of-patient-care

4. Winters-Miner, L.A.: Seven Ways Predictive Analytics Can Improve Healthcare. Elsevier, New York (2014)
5. Kornegay, C., Segal, J.B.: Selection of Data Sources. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide, pp. 109–28. Agency for Healthcare Research and Quality (US), Rockville, MD (2013)
6. Song, F., Guo, Z., Mei, D.: Feature selection using principal component analysis. In: International Conference on IEEE System Science, Engineering Design and Manufacturing Informatization (ICSEM), vol. 1, pp. 27–30 (2010)
7. Surendiran, B., Vadivel, A.: Feature selection using stepwise ANOVA discriminant analysis for mammogram mass classification. Int. J. Recent Trends Eng. Technol. 3(2), 55–57 (2010)
8. Shlens, J.: A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100 (2014)
9. Kim, H.Y.: Analysis of variance (ANOVA) comparing means of more than two groups. Restor. Dent. Endod. 39(1), 74–77 (2014)
10. Kumar, M., Rath, N.K., Swain, A., Rath, S.K.: Feature selection and classification of microarray data using MapReduce based ANOVA and K-Nearest Neighbor. Procedia Comput. Sci. 54, 301–310 (2015)
11. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, vol. 1215, pp. 487–499 (1994)
12. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. 2(1), 3 (2014)
13. Sharma, R., Singh, S.N., Khatri, S.: Medical data mining using different classification and clustering techniques: a critical survey. In: IEEE Second International Conference on Computational Intelligence & Communication Technology (CICT), pp. 687–691 (2016)
14. Yadav, C., Wang, S., Kumar, M.: An approach to improve apriori algorithm based on association rule mining. In: IEEE Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–9 (2013)
15. Ilayaraja, M., Meyyappan, T.: Efficient data mining method to predict the risk of heart diseases through frequent itemsets. Procedia Comput. Sci. 70, 586–592 (2015)
16. Rani, G.U., Prakash, R.V., Govardhan, A.: Mining multilevel association rule using pincer search algorithm. Comput. Sci. 2(5) (2013)
17. Narvekar, M., Syed, S.F.: An optimized algorithm for association rule mining using FP tree. Int. Conf. Adv. Comput. Technol. Appl. 45, 101–110 (2015)
18. Tsumoto, S.: Mining diagnostic taxonomy and diagnostic rules for multi-stage medical diagnosis from hospital clinical data. In: IEEE International Conference on Granular Computing. GRC 2007, p. 611 (2007)
19. Kaushik, S., Choudhury, A., Mallik, K., Moid, A., Dutt, V.: Applying data mining to healthcare: a study of social network of physicians and patient journeys. Machine Learning and Data Mining in Pattern Recognition. LNCS (LNAI), vol. 9729, pp. 599–613. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41920-6_47
20. Vembandasamy, K., Sasipriya, R., Deepa, E.: Heart diseases detection using Naive Bayes Algorithm. IJISET-Int. J. Innov. Sci. Eng. Technol. 2, 441–444 (2015)
21. Gulia, A., Vohra, R., Rani, P.: Liver patient classification using intelligent techniques. (IJCSIT) Int. J. Comput. Sci. Inf. Technol. 5, 5110–5115 (2014)
22. Parveen, A.N., Inbarani, H.H., Kumar, E.S.: Performance analysis of unsupervised feature selection methods. In: Computing, Communication and Applications (ICCCA), pp. 1–7. IEEE (2012)
23. Danielson, E.: Health research data for the real world: the MarketScan® Databases. Truven Health Analytics, Ann Arbor (2014)

24. KDB+ 3.4: Computer software. Kx Systems, Palo Alto (2016)
25. World Health Organization: Manual of the International Classification of Diseases, Injuries, and Causes of Death, Ninth Revision, Geneva (1977). https://simba.isr.umich.edu/restricted/docs/Mortality/icd_09_codes.pdf
26. Sayad, S.: ZeroR Classifier. http://chem-eng.utoronto.ca/~datamining/dmc/zeror.htm
27. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
28. Mitchell, T.: Decision tree learning. Mach. Learn. **414**, 52–78 (1997)
29. Witten, I., Frank, E., Hall, M.: Data Mining, pp. 102–103. Morgan Kaufmann, Burlington (2010). ISBN 978-0-12-374856-0
30. Langley, P., Sage, S.: Induction of selective Bayesian classifiers. In: Proceedings of the Tenth international Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., pp. 399–406 (1994)
31. Peng, C.Y.J., Lee, K.L., Ingersoll, G.M.: An introduction to logistic regression analysis and reporting. J. Educ. Res. **96**(1), 3–14 (2002)
32. Brownlee, J.: Logistic Regression for Machine Learning. https://machinelearningmastery.com/logistic-regression-for-machine-learning
33. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intell. Syst. Appl. **13**(4), 18–28 (1998)
34. Ting, K.M.: Precision and recall. In: Liu, L., Özsu, M. (eds.) Encyclopedia of Machine Learning, p. 781. Springer, New York (2011). https://doi.org/10.1007/978-1-4899-7993-3_5050-2
35. Dezyre: Top 10 Machine Learning Algorithms. https://www.dezyre.com/article/top-10-machine-learning-algorithms/202
36. Piatetsky-Shapiro, G.: Discovery, analysis and presentation of strong rules. In: Knowledge Discovery in Databases (1991)
37. Janecek, A., Gansterer, W., Demel, M., Ecker, G.: On the relationship between feature selection and classification accuracy. In: New Challenges for Feature Selection in Data Mining and Knowledge Discovery, pp. 90–105 (2008)
38. Motoda, H., Liu, H.: Feature selection, extraction and construction. In: Communication of IICM (Institute of Information and Computing Machinery, Taiwan), vol. 5, pp. 67–72 (2002)
39. Pearl, J.: Entropy, information and rational decisions. Technical report. Cognitive Systems Laboratory, University of California, Los Angeles (1978)
40. Russell, S., Norvig, P.: Artificial Intelligence. A modern approach, vol. 25, p. 27. Prentice-Hall, Egnlewood Cliffs (1995)
41. Bayes, M., Price, M.: An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. Philos. Trans. (1683–1775) **53**, 370–418 (1963)
42. Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. Stat. Comput. **14**(3), 199–222 (2004)
43. Wickens, T.D.: Elementary Signal Detection Theory. Oxford University Press, Oxford (2002)
44. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y.: Artificial intelligence in healthcare: past, present and future. Stroke Vasc. Neurol. SVN **2**, 230–243 (2017)
45. Rajeswari, K., Vaithiyanathan, V., Pede, S.V.: Feature selection for classification in medical data mining. Int. J. Emerg. Trends Technol. Comput. Sci. (IJETTCS) **2**(2), 492–497 (2013)