Predicting colorectal cancer with the aid of temporal patterns

Jullian van Kampen

VU university, Amsterdam
Master Business Analytics
De Boelelaan 1081
1081 HV Amsterdam

Abstract

A patient suffering from colorectal cancer may not realize it is suffering from this disease until it is too late. Therefore it is important to discover this disease as soon as possible since this will increase the survivability of the disease. Electronic Medical Records (EMR) databases contain information about a patient's medical history which can help in predicting whether a patient has colorectal cancer. However, when a predictive model is generated, time aspects of these visits are often not used. In this paper an attempt is made to use this time aspect to create and exploit temporal patterns for predictive modelling. A temporal pattern is a sequence of events which share a relation to each other. In order to create these patterns an a-priori mining algorithm was used which is then further exploited by considering the time needed to complete such a pattern. The added value of these patterns are tested using logistic regression, a decision tree, Support vector machine, and random forest. Each algorithm is executed twice: one time with the patterns generated by the a-priori algorithm and another time with the patterns generated by the exploitation scheme. As a result, The performance of the models improved by using temporal patterns in their predictive model. This shows that temporal data potentially holds vital information for practitioners in the task of identifying colorectal cancer

Introduction

An early detection of colorectal cancer can be of importance for the survival of a patient. Although a lot of research has been performed, colorectal cancer still is the third most common form of cancer in both man and women [1]. Methods like screening and early intervention can help in an early detection of colorectal cancer and reducing the mortality rate of colorectal cancer.

Temporal data mining is used to find relations in sequences of events which are not obvious at first sight. [2] describes three steps in finding these hidden relations and techniques that can be used to find them. The first step covers the representation and preprocessing of the sequences for the actual data mining operations. The second step is defining a similarity measure to see whether sequences match and occur frequently. The last step covers the actual mining of the sequences using a mining algorithm.

According to [3] The algorithms that are used to mine these patterns can be divided into 3 categories namely apriori-based, pattern-growth and early-pruning. Moreover, the investigation of these categories reveals that certain heuristics are required for a reliable sequential pattern mining algorithm. There are however also approximate pattern mining algorithms which allow a certain degree of error in their discovered patterns [4].

Previous research has already shown that some symptoms can be related with colorectal cancer (CRC). [5] gives an overview on which symptoms are known to predict CRC. The most common symptom was either rectal bleeding by itself or rectal bleeding combined with anaemia, constipation, or abdominal pain. In a different study [6] shows that patients with constipation have an increased risk of having CRC. Moreover, the study showed that the use of laxatives increases the probability of getting CRC.

In this paper temporal data mining is used to discover whether temporal data can contribute to the prediction of CRC. Many researchers have tried to predict CRC with the use of temporal data and modelling techniques. [7] shows that the model performance improves if certain events that occur for a patient are stored in a temporal pattern. A temporal pattern is a set of events in succession which are then described by a relation between these events. The algorithm used to find these patterns however ignores the time aspect of the events which could result in better performance. Therefore, an attempt is made to further improve these patterns by adding information about the time it takes to complete a temporal pattern.

First a description of the dataset is given. Next insight is provided in how the data was prepared for the algorithms which also describes the algorithms used to mine temporal patterns. After that the methodology of our research is discussed which contains our research questions, the algorithms used and information on how the pattern completion times are added. This is followed by the results of each of these algorithms where the influence of pattern completion times are evaluated. Finally, we conclude this paper with a discussion on the results of the algorithms, answers to the research questions and possible future work.

Data description

The dataset that was made available consists of anonymized data from the Utrecht region in the Netherlands between 01-07-2006 and 31-12-2011. This dataset contains information about 219.447 patients who have visited a general practitioner which registers each visit with a certain code. Therefore, the data is provided in 4 different parts. The first part of the dataset covers general patient information. Each patient has a unique ID and information about their gender, date of birth and their register date. The second part of the data provides information on the consults of a patient. Every time a patient consults a general practitioner an ICPC (International Classification of Primary Care) code is provided before and after consult. This code is used to describe diseases and symptoms of a patient. Figure 1 shows that the amount of consults seems to peak during November and is at its lowest in june and juli.

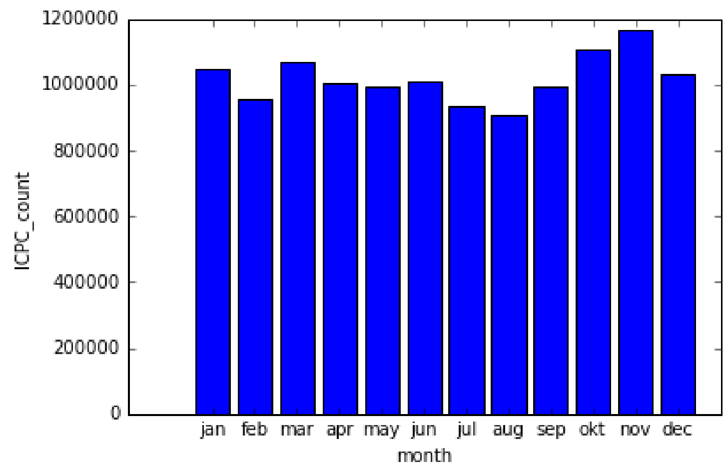


FIGURE 1: aggregate amount of consults for each month over the years 2007-2011

The third part of the data contains information on the medication of a patient. Once a patient receives medication, an ATC (Anatomic Therapeutical Chemical) code is provided describing the medication received. Figure 2 shows that there seems to be a high amount of drug prescriptions in the December month and a lower amount of prescription in the January and February month. Finally, the last part of the data contains information on the referrals of a patient to a specialist. Figure 3 shows that most of the referrals occur at the start of the year and that later on the general practitioner gives less referrals.

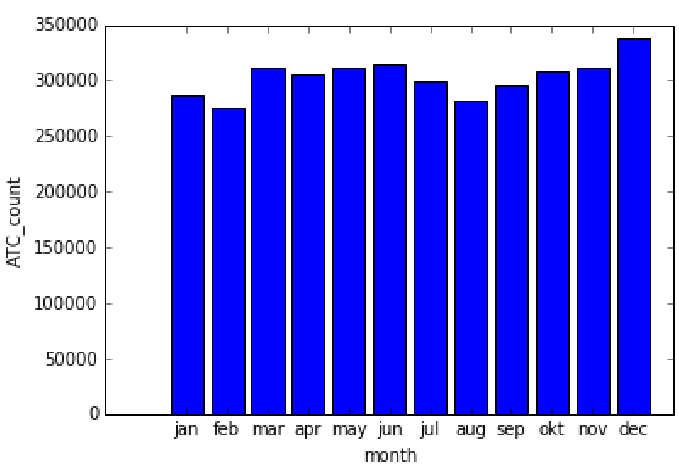


FIGURE 2: aggregate amount of drug prescriptions for each month over the years 2007-2011

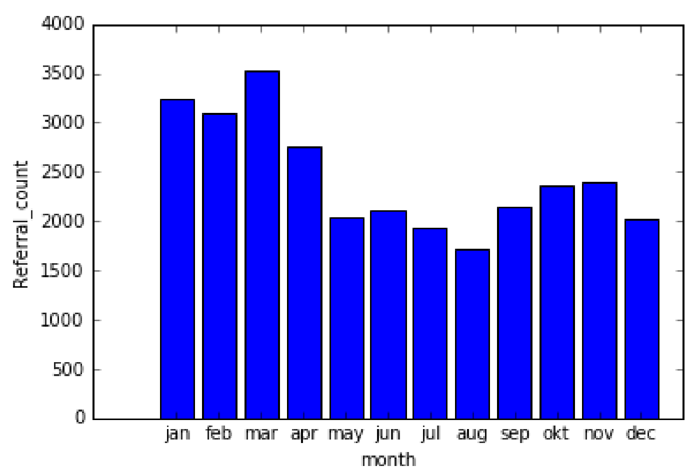


FIGURE 3: aggregate amount of referrals for each month over the years 2007-2011

Data preparation

Before datamining algorithms can be applied the data needs to be prepared. Therefore the dataset goes through a pipeline which covers the preprocessing of the data.

Pipeline

First of all, the pipeline goes through the patient files and records the patient's registration date, age and gender. If the age of a person is lower than 30 years old the patient is removed from the dataset. This is done since CRC is considered more relevant for people with a higher age. Next, all the patients which are diagnosed with CRC are looked up. This is diagnosed in a patient with ICPC code D75. From there the dataset is split in two parts: people diagnosed with CRC, and people diagnosed without CRC. This will be important later on since we want to know which patterns occur mostly in CRC patients and which don't. This resulted in 808 known CRC cases in the dataset. Once the CRC patients are known, a time period has to be chosen from which a sample is taken. This has to be done since the dataset is too large to process all at once. Therefore, a 6-month interval is taken of every patient to be used for the preprocessing. For CRC patients the interval was taken over the 6 months before they were diagnosed with CRC. For non-CRC patients, a random 6 month interval was taken. If for some reason an interval could not be generated, the patient is removed from the dataset.

Finally, the pipeline goes through the remaining files to extract the standard attributes. These standard attributes consists of a patient's age, gender and the events that occur over the 6-month time interval. These events can be either an ATC code, ICPC code, or a referral to a specialist. Eventually, for each patient it is known which events took place and how many times it took place for that specific patient. Once all the standard attributes are generated, we can start working on generating temporal patterns. These patterns were found by using a temporal pattern mining approach borrowed from [7]. The basic idea of the algorithm is explained below.

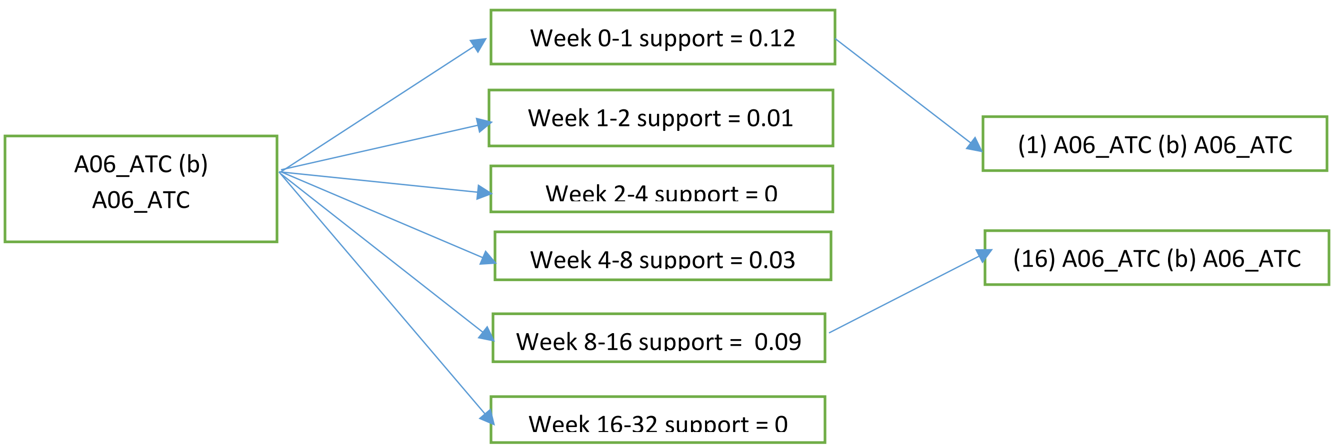
Temporal patterns

First, all the records are scanned and frequent 1-patterns [1] are created. A pattern is considered frequent if the pattern is found in a certain minimum percentage of patients. This percentage is referred to as the support of the pattern. Then, these frequent 1-patterns are used to create frequent 2-patterns. Thus, once the frequent k-patterns have been generated, frequent k+1-patterns are created. This process continues until no more frequent patterns can be found. Every k+1-pattern also describes a relation between the events. This relation can be either co-occurrence (c), meaning they happen at the same date, or succession (b), which means that one event precedes the other. A full explanation on how to generate these frequent patterns and optimizations on this approach can be read in [8]. Finally, the algorithm is run on CRC cases and on non-CRC cases separately so that predictive patterns can be found for both classes.

[1] A 1-pattern describes the presence of a certain attribute in a patient while a 2-pattern describes the presence of 2 attributes in a certain order with a certain relation

Pattern completion time

Once the temporal datamining algorithm generated the frequent patterns, the so called pattern completion time can be computed. The pattern completion time tells us whether a k+1-pattern is completed within 1,2,4,8,16 or 32 weeks. As a result a frequent pattern can occur more than once but including information about its pattern completion time. The flow diagram below illustrates this procedure. Assume only patterns with a minimum support of 0.05 or higher are taken. In the first block a pattern was found with a support of 0.25 so it is frequent. Next, for each patient is checked how much time was needed to complete this pattern in weeks. As described before there are 6 possible intervals in which the pattern could be completed. For each interval it is checked whether each of these new support values is larger than the minimum support. In our example this is the case for the weeks 0-1 and 8-16. As a result 2 additional patterns were found which hold information on the pattern completion time in addition to the complete pattern without this information.



Methodology

The purpose of this paper is to investigate whether temporal patterns can help to identify colorectal cancer. Therefore, an attempt is made to confirm the following hypothesis. Can time aspects in temporal data contribute to predicting colorectal cancer for a patient? In order to answer this hypothesis, an attempt is made to answer the following research questions:

1. Which pattern completion times are more common and which are less common and which ATC and ICPC codes are used to generate these patterns
2. Does the pattern completion time help the algorithm in predicting CRC
3. Do patterns with lower completion time have more predictive value than patterns with a high completion time.

For the first research question the temporal pattern mining algorithm was used with different support values and checked what kind of pattern completion times are more common than others. Additionally a count is given for the codes belonging to a certain chapter to see which codes appear more frequently over all patterns. For the second part the results of the dataset excluding pattern completion time is compared to the dataset including pattern completion time. The results will be tested by applying four machine learning algorithms to each dataset. Finally, the results of these algorithms were investigated to see if the algorithms had any preference for patterns with a low completion time or patterns with a high completion time.

Algorithms

For the algorithms that were used we decided to use the same setup that was used in [7]. This was done so that we are able to use this reference as a benchmark for our results. Moreover, using the same setup as in [7] means that we also don't have to do anything about parameter tuning which is beyond the scope of this paper.

The machine learning algorithms used are all implemented in the Python Scikit-Learn package. first a CART(Classification And Regression Tree) [9] decision tree was used. The parameters for this tree are a maximum depth of 5, The minimum number of samples on a leaf node is 50 and the splitting criteria is based on the gini impurity measure. Next a Random Forest (RF) [10] algorithm was used on the data. The same parameter settings were used as with CART with the addition of the size of the forest set to 100. After RF a logistic regression (LR) [11] model was fitted to the data. As a regularization parameter the default value L2 was chosen. Finally a Support Vector Machine (SVM) [12] was used with a RBF kernel type

Results

Temporal patterns

First the temporal mining algorithm was executed with a minimum support of 0.10. This means that 10% of the patients should cover the pattern in order for it to be frequent. As a result 71 patterns were found of which 27 contained information about the pattern completion time. A complete list of the created patterns can be found in the appendix. Figure 4 shows that most of the patterns are completed either within the first week or somewhere between week 8 and 16. This can be explained by the fact that most of the patterns found are from patients with CRC and thus visit a practitioner or receive a drug prescription on short notice. The longer interval can be explained by patients which have chronical diseases and thus receive the same drug prescription after a certain amount of months. Below two bar charts can be found displaying which ICPC and ATC codes were most commonly used to build patterns. The horizontal axis shows what chapter the code belongs to which is the first letter of the code. For example, in the code D12_ICPC the D

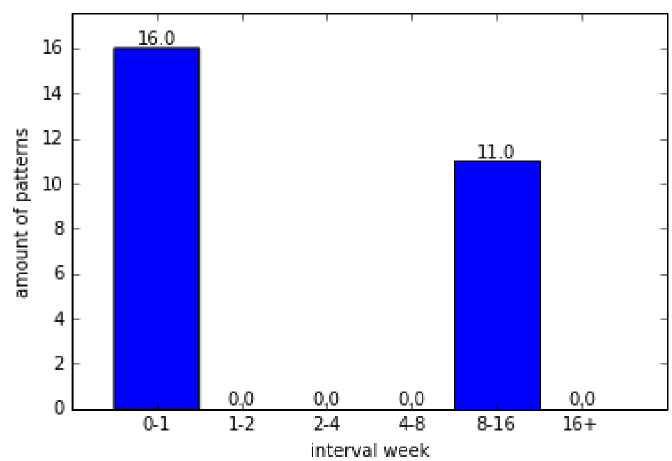


Figure 4: pattern completion times with minimum support of 0.10

describes that the patient has complaints considering their digestive system. Overall, the algorithm seems to find more patterns containing ATC codes than ICPC codes. Furthermore, figure 6 shows that the ICPC code most commonly used in the patterns is K which describes circulatory complaints. Figure 5 shows that by far the most common used ATC code in the temporal patterns is C which describes medication for the cardiovascular system.

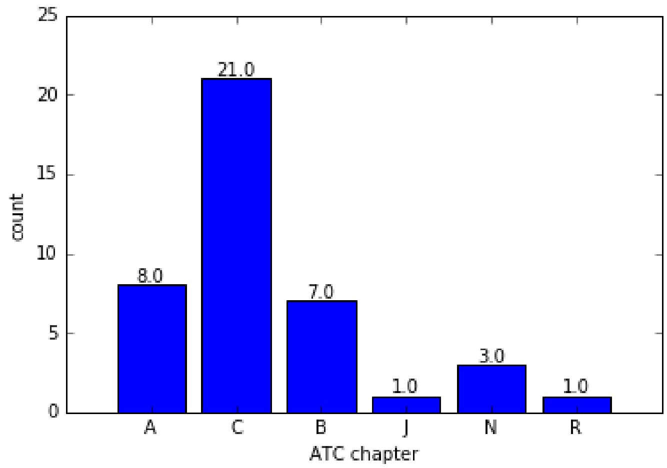


FIGURE 5: ATC codes used to create patterns

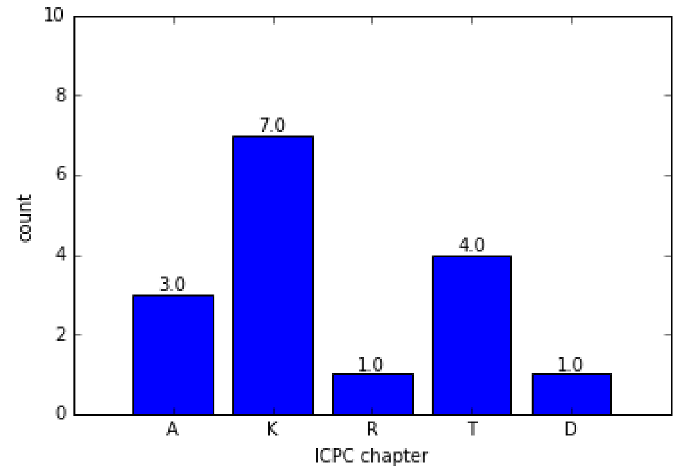


FIGURE 6 : ICPC codes used to create patterns

Increasing the minimum support would not make any sense, since it would only result in less patterns for the same weeks. Moreover, we are interested in the effect of temporal patterns on the machine learning algorithms, so we want a lot of patterns. Therefore, the algorithm was executed once more but now the minimum support was lowered to 0.05. Lowering the support even further would probably result in even more patterns and perhaps a different distribution. This was not done however since the quality of the patterns would suffer from a low support value. This resulted in 467 patterns of which 236 contained information about their pattern completion time. Once again a complete list of created patterns can be found in the appendix. Figure 7 shows that the results are almost similar to the previous result with a support of 0.10.

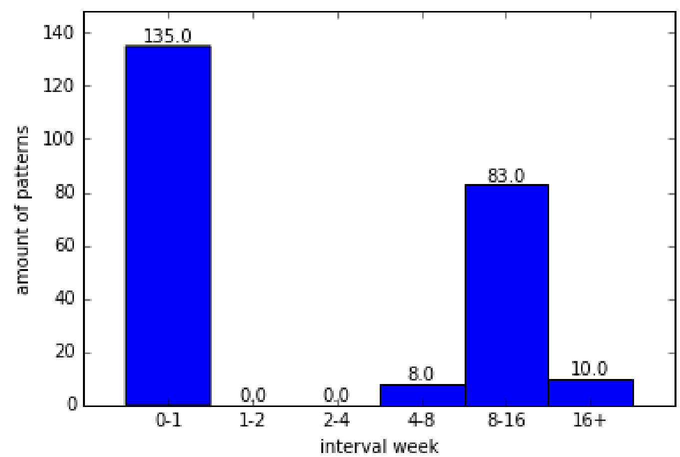


FIGURE 7: pattern completion times with minimum support of 0.05

If we once again look at the codes which are most commonly used, it can be seen from figure 9 that still most of the ICPC codes contain the chapters K, A and T. Some other codes are introduced as well due to the lower support value but they do not seem to be of any importance. Figure 8 shows that the ATC codes A, B and C are most commonly used while the other codes are more or less neglected.

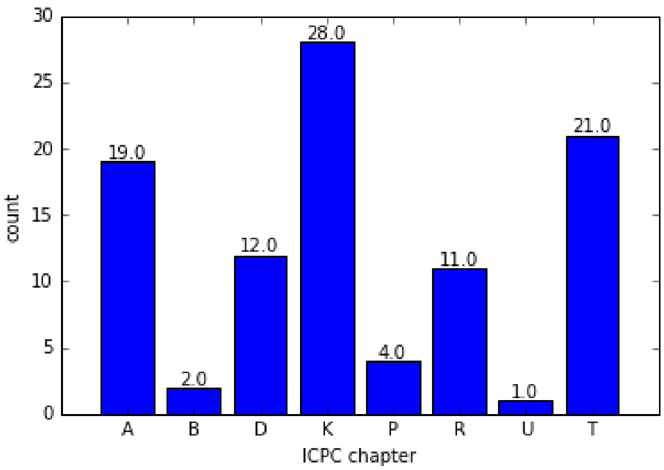


FIGURE 9: ICPC codes used to create frequent patterns

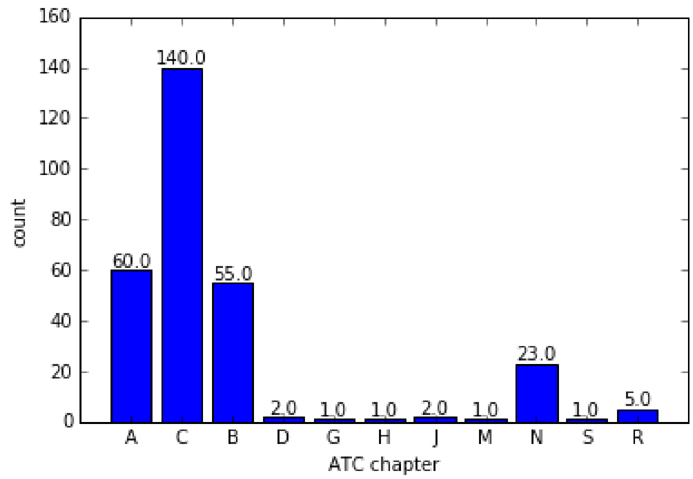


FIGURE 8 : ATC codes used to create frequent patterns

Next the results of the algorithms on each dataset are compared and investigated to see which variables are used to come to this result. Before the algorithm starts it selects 50 attributes which have the lowest pearson correlation. The performance of the model is evaluated by the AUC (Area Under the Curve) and with the use of confusion matrices which is acquired by applying a 5-fold cross validation scheme. For the AUC a 95% confidence interval is provided to investigate whether the algorithm performs significantly better. The computation of this interval is based on [13]. In order to compare the confusion matrices a constant false positive rate of 0.40 on the AUC curve was chosen for the dataset with and without pattern completion times. Based on this false positive rate a threshold is computed for the dataset with and without pattern completion times which tells us what cutoff should be used to classify a person as a CRC patient or a non-CRC patient. Additionally, our main interest is in the use of the temporal patterns for predicting CRC. Therefore, the same analysis for each algorithm were performed with a minimum support of 0.05 to see whether the machine learning algorithms prefers the patterns over the usual variables.

CART

Using the CART algorithm on the dataset without pattern completion time resulted in a decision tree where not a lot of patterns were used. Table 1 shows that the only pattern that was used is the 1-pattern A06_ATC, which describes a drug for constipation. This split however could have also been made with the standard attribute a06_ATC. Moreover, the gini coefficient is rather low so it seems that the patterns do not have any influence on the current decision tree. The rest of the tree is made up of the known predictors rectal bleeding (d16_ICPC) iron deficiency aneamia (b80_icpc), change in bowel movements (d18_icpc) abdominal pain (d01_icpc) constipation (d12_icpc) and the age of a patient. The full decision tree can be found in the appendix of this report.

Pattern	Description	Depth	Gini coefficient
8	A06_ATC	1	0.012

TABLE 1: patterns used by CART with 0.10 support excluding pattern completion times

If pattern completion times are added it is observed that there are several more patterns in the decision tree, and that the decision tree is wider. Surprisingly the decision tree favors 2-patterns over the previously found 1-patterns, and most of them are drugs with chapter code A. Furthermore, the algorithm seems to take advantage of the pattern completion time since it first checks whether pattern 24 is present and after that if pattern 23 is as well. Table 2 shows the patterns that were used in the decision tree and at what depth. A02_ATC describes drugs for acid related disorders and A06_atc describes drugs for constipation. The rest of the decision tree looks almost the same as the previous except that in this tree d12_icpc has been removed. In return b03_atc which describes antianemic drugs and a 3-pattern describing uncomplicated hypertension (K86_icpc) were added. Once again the full decision tree can be found in the appendix

Completion time in weeks	Pattern	Description	Depth	Gini coefficient
0-32	4	A02_ATC (b) A02_ATC	4	0.0638
0-32	9	A02_ATC (b) A06_ATC	3	0.2563
0-1	23	A06_ATC (b) A06_ATC	1	0.1902
0-32	24	A06_ATC (b) A06_ATC	0 (root)	0.0128
0-32	59	K86_ICPC (b) K86_ICPC (b) K86_ICPC	3	0.08

TABLE 2: patterns used by CART with 0.10 support and including pattern completion times

The threshold value used to classify a patient as a CRC patient decreased from 0.2599 to 0.1957. Addition of the pattern completion time resulted in more false negatives and less true positives. Additionally more true positives and less true negatives were found. In the decision tree without pattern completion times an AUC of 0.8438 is acquired with a confidence interval of (0.8240-0.8637). In the decision tree with pattern completion times an AUC of 0.8671 was found and a confidence interval of (0.8485 – 0.8859). Since the confidence intervals overlap, the performance of the model is not significantly better. The F-measure decreases from 0.0308 to 0.0305 since the decrease in recall outweighs the increase in recall. These scores may seem rather low but the main interest is the comparison of the score to each other.

	Pred 0	Pred 1
Actual 0	56379	33968
Actual 1	48	540

TABLE 4: confusion matrix for CART without pattern completion times

	Pred 0	Pred 1
Actual 0	55309	35038
Actual 1	36	552

TABLE 3: confusion matrix for CART with pattern completion times

When the same algorithm is run with a support of 0.05 similar results for the dataset without pattern completion times are obtained. The decision tree obtained looks exactly the same as the one previously found. If the pattern completion times are added the decision tree changes drastically. Of the known predictors, only the age variable can be found in the decision tree at different depths. These known predictors however can be retrieved from the patterns that were generated. As with the previous results, the decision tree uses more and longer patterns. Once again the algorithm seems to take advantage of the pattern completion times since it includes the pattern A06_ATC (b) A06_ATC without a completion time of 0-32 weeks and A06_ATC (b) A06_ATC which includes a completion time over 0-1 weeks[2]. The table below shows which patterns were used by the decision tree and that some new atc codes are introduced in the tree. B01_atc describes antithrombotic agents, where C07_atc describes beta blocking agents and C09_atc describes agents acting on the renin-angiotensin system.

Completion time in weeks	Pattern	Description	Depth	Gini coefficient
0-32	337	D16_ICPC (b) D16_ICPC	0	0.0128
0-32	320	A06_ATC (b) A06_ATC	1	0.0117
0-1	319	A06_ATC (b) A06_ATC	2	0.1719
0-32	298	B01_ATC (b) B01_ATC (b) B01_ATC	2	0.0084
0-32	228	K86_ICPC (b) K86_ICPC (b) K86_ICPC	3	0.0839
0-32	138	D18_ICPC	3	0.0069
0-32	187	B80_ICPC (b) B80 ICPC	4	0.0065
0-32	204	C07_ATC (b) C07_ATC (b) C07_ATC	4	0.0316
0-32	178	C09_ATC (b) C09_ATC (b) C09_ATC	4	0.1204
0-32	140	D01_ICPC	4	0.1686

TABLE 5: patterns used by CART with 0.05 support excluding pattern completion times

Without the pattern completion times the algorithm scored an AUC of 0.8485 with a confidence interval of (0.8288 – 0.8681). When pattern completion times were added, the AUC score increased to 0.8779 with a confidence interval of (0.8598 – 0.8960). There is a slight overlap of the confidence intervals so therefore it is hard to say whether the algorithm performs significantly better when pattern completion times are added. The threshold value used to classify a patient as a CRC patient decreased from 0.2594 to 0.2162. An improvement is found in the amount of true negatives and a decrease in the amount of true positives. The recall and the precision both decrease slightly which results in a deterioration in the F-measure from 0.0296 to 0.0295.

	Pred 0	Pred 1
Actual 0	54305	36042
Actual 1	38	550

TABLE 6: confusion matrix for CART without pattern completion times

	Pred 0	Pred 1
Actual 0	56709	33638
Actual 1	40	548

TABLE 7: confusion matrix for CART with pattern completion times

[2] Pattern numbering may change when the algorithm is run with different support.

LR

The table below shows the 10 best predictors for the logistic algorithm with and without using pattern completion times in the data. Excluding pattern completion time some new codes are used which previously were not used. R44_ICPC and T46_ICPC are codes of which we do not known what they imply. The addition of pattern completion times changes the variables selected to predict CRC. This is possibly caused by the introduction of p23 which has a pattern completion time included. Moreover one new ICPC and ATC code are introduced in the best 10 variables. A99_ICPC describes a class of general diseases and C03_ATC describes diuretic drugs. Additionally, just like the decision tree algorithm, logistic regression seems to prefer 2-patterns over 1-patterns once completion times are introduced

Position	Best 10 excluding pattern completion time	Best 10 including pattern completion time
1	d11_ICPC	d16_ICPC
2	Age	b03_ATC
3	t46_ICPC	C03_ATC (b) C03_ATC (p18)
4	A02_ATC (b) A06_ATC (p7)	A99_ICPC (b) A99_ICPC (p14)
5	A06_ATC (p8)	A06_ATC (b) A06_ATC (p23) (0-1)
6	B03_ATC (p31)	B03_ATC (p44)
7	B01_ATC (p40)	D12_ICPC (p65)
8	R44_ICPC (p15)	d11_ICPC
9	d12_ICPC	A02_ATC (b) A06_ATC (p9)
10	d18_ICPC	t46_ICPC

TABLE 8: Features found using Logistic Regression with 0.10 support

The algorithm without pattern completion times gives an AUC of 0.8883 and a confidence interval of (0.8709-0.9058). The algorithm with pattern completion times gives an AUC of 0.9271 and a confidence interval of (0.9125-0.9417). Since the confidence intervals do not overlap, the algorithm seems to perform significantly better with the addition of the pattern completion times. The threshold value used to classify a person as a CRC patient decreased from 0.0031 to 0.0025. Furthermore, they show that more true positives and more true negatives are found. Since both the precision and the recall increase, The F1-measure shows an improvement from 0.0299 to 0.0307 when pattern completion times are included.

	Pred 0	Pred 1
Actual 0	54213	36134
Actual 1	30	558

TABLE 9: Confusion matrix for Logistic Regression without pattern completion times

	Pred 0	Pred 1
Actual 0	54429	35918
Actual 1	18	570

TABLE 10: Confusion matrix for Logistic Regression with pattern completion times

When additional patterns are introduced by lowering the support needed to 0.05, the algorithm seems to select different variables once again to predict CRC. Although the standard variables remain unchanged in the top 10 the 1-patterns seem to go away and make room for 2-patterns. These 2-patterns however contain predictors that were observed before, like drugs for constipation (A06_ATC) and rectal bleeding (D16_ICPC). Additionally a new variable enters the top 10 namely b82_icpc which describes unspecified anaemia. If the algorithm uses the dataset including pattern completion times it seems to mostly use 2-patterns and 3-patterns to predict CRC. Once again the pattern containing rectal bleeding (D16_ICPC (b) D16_ICPC) is chosen as one of the most predictive patterns. This time however a pattern completion time is included Moreover, the algorithm uses the patterns containing pattern completion times to predict CRC. The 2-patterns and 3-patterns used in the algorithm also contain predictors which were observed before.

Position	Best 10 excluding pattern completion time	Best 10 including pattern completion time
1	d11_ICPC	p336 (D16_ICPC (b) D16_ICPC) (0-1)
2	d16_ICPC	d18_ICPC
3	p102 (D16_ICPC)	p298 (B01_ATC (b) B01_ATC (b) B01_ATC)
4	Age	Age
5	p37 (C09_ATC (b) A06_ATC)	b80_ICPC
6	b82_ICPC	p320 (A06_ATC (b) A06_ATC)
7	b80_ICPC	p142 (D01_ICPC (b) D01_ICPC)
8	p161 (A06_ATC (b) A06_ATC)	p42 (A02_ATC (b) A02_ATC (b) A02_ATC) (4-8)
9	t46_ICPC	p178 (C09_ATC (b) C09_ATC (b)C09_ATC)
10	p143 (B80_ICPC)	a06_ATC

TABLE 11: Best features found using Logistic Regression with 0.05 support. The numbers at the end of a pattern specify the interval in which the pattern occurs

Without the pattern completion times The model scored an AUC of 0.8933 with a confidence interval of (0.8761 – 0.9105). When pattern completion times were added, the performance of the model increased to 0.9292 with a confidence interval of (0.9148 – 0.9436). The confidence interval don't overlap so once again the logistic regression algorithm performs significantly better with pattern completion times. The threshold value used to classify a patient as a CRC patient decreased from 0.0029 to 0.0023. The algorithm is able to find more true negatives and also finds more true positives. Thus, the quality of the model improves as well as the performance of the model. Since both precision and recall increase the F-measure shows an increase from 0.0299 to 0.0305.

	Pred 0	Pred 1
Actual 0	54244	36103
Actual 1	30	558

TABLE 13: Confusion matrix for Logistic Regression without pattern completion times

	Pred 0	Pred 1
Actual 0	54299	36048
Actual 1	20	568

TABLE 12: Confusion matrix for Logistic Regression with pattern completion times

RF

The table below shows the 10 most predictive variables for the random forest algorithm. In the RF without pattern completion times we don't see a lot of change in the predictive variables compared to LR. It does however include the referral to a specialist. The pattern completion time seems to have had some influence on the chosen attributes although they are not present in the top 10. As with the previous algorithms d16_ICPC scores well compared to how it performed without pattern completion times. Finally the patterns found in the original temporal mining algorithm do not hold any significant value in the new algorithm.

position	RF without pattern completion time	Position in RF+	RF with pattern completion time (RF+)	Position in RF
1	d18_ICPC	2	d16_ICPC	50+
2	b80_ICPC	3	d18_ICPC	1
3	A06_ATC (b) A06_ATC (p17)	50+	b80_ICPC	2
4	Age	5	A02_ATC (b) A02_ATC (p4)	40
5	a06_ATC	8	Age	4
6	A06_ATC (p8)	7	d01_ICPC	10
7	gastro-enterologie_specialisme	9	A06_ATC (p10)	6
8	b03_ATC	19	a06_ATC	5
9	b82_ICPC	14	gastro-enterologie_specialisme	7
10	d01_ICPC	6	A99_ICPC (b) A99_ICPC(p14)	38

TABLE 14: Best features found using Random Forest with support 0.10

Without pattern completion times the AUC score was 0.8850 and a confidence interval of (0.8672-0.9026). With pattern completion times the AUC score was 0.8912 and the confidence interval was (0.8740-0.9085). Once again the confidence interval overlap and therefore the results are not significant. The threshold value for classifying CRC patients decreased from 0.0030 to 0.0027. Results show that the amount of true positives increases and that the amount of true negatives increases. This means that the quality of the solution improves when pattern completion times are added. Both the precision and the recall increases which means that the F1-measure increases from 0.0295 to 0.0299.

	Pred 0	Pred 1
Actual 0	54209	36138
Actual 1	38	550

TABLE 15: Confusion matrix for Logistic Regression without pattern completion times

	Pred 0	Pred 1
Actual 0	54345	36002
Actual 1	33	555

TABLE 16: Confusion matrix for Logistic Regression with pattern completion times

If the algorithm is rerun with the patterns that were obtained with a support of 0.05, RF did take more patterns in its top predictors. Most of these patterns are 1-patterns which can also be found in the top 10 as a standard attribute. The forest however does seem to use a smaller variety of codes in the top 10 than in the previous case. When pattern completion times are added, RF adds two variables in the top 10 with information about their pattern completion time. Once again the pattern describing rectal bleeding (D16_ICPC (b) D16_ICPC) seems to be the most predictive pattern. Additionally, the top 10 now mostly consists out of 2 and 3-patterns of ATC and ICPC codes seen before.

position	RF without pattern completion times	Ranking in RF+	RF with pattern completion time (RF+)	Ranking in RF
1	d18_ICPC	13	P336 (D16_ICPC (b) D16_ICPC) (0-1)	-
2	p71 (D18_ICPC)	4	p320 (A06_ATC (b) A06_ATC)	11
3	d16_ICPC	9	p228 (A06_ATC (b) A06_ATC (b) A06_ATC)	39
4	p102 (D16_ICPC)	10	p138 (D18_ICPC)	2
5	p97 (B80_ICPC (b) B80_ICPC)	15	p186 (B80_ICPC (b) B80_ICPC)(0-1)	-
6	Age	7	d01_ICPC	15
7	p226 (A06_ATC)	16	Age	6
8	a06_ATC	23	p142 (D01_ICPC (b) D01_ICPC)	14
9	b80_ICPC	14	d16_ICPC	3
10	p143 (B80_ICPC)	32	p196 (D16_ICPC)	4

Table 17: Best features found using Random Forest with 0.05 support

When the algorithm is rerun with a minimum support of 0.05, the performance of the model does not improve much. Without pattern completion times the model was able to score an AUC of 0.8855 with a confidence interval of (0.8678 – 0.9031). When the pattern completion times are added the AUC increased to 0.9007 (0.8841 – 0.9174). Clearly the results are not significant since the confidence intervals don't overlap. The threshold value used to classify a patient as a CRC patient decreases from 0.0030 to 0.0028. Furthermore, results show that more true negatives are found and more true positives when adding pattern completion times. The F-measure shows an increase from 0.0298 to 0.0307 due to both a higher recall and precision.

	Pred 0	Pred 1
Actual 0	54292	36055
Actual 1	34	554

Table 18: Confusion matrix for Random Forest without pattern completion times

	Pred 0	Pred 1
Actual 0	54707	35640
Actual 1	23	565

Table 19: Confusion matrix for Random Forest with pattern completion times

SVM

For support vector machines the influence of the variables on the model can't be analyzed. however the confusion matrices and the AUC that were found by the model can be analyzed. Without pattern completion times the model gave an AUC score of 0.8199 and a confidence interval of (0.7990-0.8407). With pattern completion times the model gave an AUC score of 0.8322 and a confidence interval of (0.8118-0.8526). As a result the performance of the model is not significantly better. The threshold value used to classify a person as a CRC patient decreases from 0.0019 to 0.0018. Furthermore, results show that more true negatives and less true positives were found. As a result Both recall and precision slightly decrease which also decreases the F1-measure from 0.0296 to 0.0294.

	Pred 0	Pred 1
Actual 0	54270	36077
Actual 1	38	550

TABLE 20: Confusion matrix for Support Vector Machine without pattern completion

	Pred 0	Pred 1
Actual 0	54366	35981
Actual 1	41	547

TABLE 21: Confusion matrix for Support Vector Machine with pattern completion

If the algorithm is rerun with a support of 0.05 the model without pattern completion times gave an AUC of 0.8145 and a confidence interval of (0.7934 – 0.8356). When pattern completion times were added the AUC increased to 0.8298 with a confidence interval of (0.8093-0.8503). The confidence intervals are overlapping and therefore the results are not significantly better. The threshold value used to classify a patient as a CRC patient increases from 0.0017 to 0.0019. They show that more true positives are found but also less true negatives. Thus, with pattern completion times the model is able to find more CRC patients but at the cost of healthy patients being classified as CRC patients. Both precision and recall increases, and as a result the F-measure increases from 0.0292 to 0.0300.

	Pred 0	Pred 1
Actual 0	54464	35883
Actual 1	47	541

TABLE 22: Confusion matrix for Support Vector Machine without pattern completion

	Pred 0	Pred 1
Actual 0	54341	36006
Actual 1	31	557

TABLE 23: Confusion matrix for Support Vector Machine with pattern completion

Conclusion

Finally, a summary of the performance of the model is given to see which algorithm had the most benefit of the pattern completion times. The table below shows the AUC scores of all the algorithms and their respective confidence intervals. Additionally, the results of [7] are included to use as a benchmark which used a minimum support of 0.1. The only difference is that in [7] comorbidity data and lab results are included in the dataset. The table shows that overall the addition of pattern completion times has improved the results of the temporal mining algorithm. Moreover the AUC scores in bold show that these datasets performed significantly better than the benchmarked performance. LR and RF seem to benefit the most from the pattern completion times while RF and SVM don’t.

support	CART	LR	RF	SVM
0.1	0.844 (0.824-0.864)	**0.888 (0.871-0.906)**	0.885 (0.867-0.903)	0.820 (0.800-0.841)
0.1+	**0.866 (0.849-0.886)**	**0.927 (0.913-0.942)**	0.891 (0.874-0.909)	0.832 (0.812-0.853)
0.05	0.849 (0.829-0.868)	**0.893 (0.876-0.910)**	0.885 (0.868-0.903)	0.815 (0.793-0.836)
0.05+	**0.878 (0.860-0.896)**	**0.929 (0.915-0.944)**	0.901 (0.884-0.917)	0.830 (0.809-0.850)
benchmark	0.818 (0.798-0.838)	0.796 (0.775-0.817)	0.881 (0.864-0.898)	0.832 (0.813-0.851)

TABLE 24: Summary of the performance of the models and their confidence interval given a certain support. The + sign indicates that pattern completion times were included. Bold AUC scores indicate that it performed significantly better than the benchmark dataset

Discussion

This study was performed to investigate whether the use of temporal patterns could be exploited further for the prediction of CRC. This was done by generating patterns with a known algorithm which has proven to be successful and adding information about the time required to complete such a pattern. As a result, it seemed most of the patterns found were completed within either 1 week or between 8-16 weeks. Moreover, CART, LR and RF seemed to be using more and longer patterns in their predictions for CRC with the addition of pattern completion times for the temporal patterns. These patterns contained mainly alimentary tract and metabolisma related drug prescriptions (A##_ATC) or ICPC codes describing digestive complications (D##_ICPC). Some of these codes (A06_ATC and D16_ICPC) have been found to be predictive for colorectal cancer [6] [5]. Additionally, the temporal patterns with completion time of at most 1 week were most commonly used by the algorithms.

Overall the results show an improvement in performance of the model, and an improvement in the quality of the solution. The confusion matrices show that most of the times we are able to identify more CRC patients with the addition of pattern completion times although sometimes at the cost of less true negatives. The only exceptions are SVM with a minimum support of 0.10 and CART with a minimum support of 0.05. Furthermore, the patterns used for predicting CRC become longer and some of those longer patterns don't hold any information on the completion time since the algorithm could not find enough support. These longer patterns also affects the quality of the solution since it may take more time for a patient to create a 2 or 3 pattern which predicts CRC in a patient. Finally, we conclude that time aspects of temporal data can contribute to the prediction of colorectal cancer but the quality of the solution might suffer from it.

For future work it would be interesting to further improve the quality of the patterns. This can be done for example by forcing a combination of acid related ATC codes with digestive ICPC codes in a pattern. This might give more insight in the reaction of a patient to a certain drug prescription. Furthermore, the patterns found in this study need to be validated by other studies to see whether the results found are incidental or not. Finally, a pattern can have more relations then only co-occurency and succession as described in [8].

References

1 Siegel, R., DeSantis, C., Jemal, A. Colorectal Cancer Statistics, 2014. *CA: A Cancer Journal for Clinicians*, 64, 2 (march/april 2014), 104-117.

2 Antunes, C.M., Oliviera A.L. *Temporal Data Mining: an overview*. instituto Superior Técnico, Lisboa, Portugal, 2001.

3 Mabroukeh, N. R., Ezeife C. I. A Taxonomy of Sequential Pattern Mining Algorithms. In *ACM computing surveys 43* (USA 2010).

4 Gupta, R., Fang, G., Field, B., Steinbach, M., Kumar, V. Quantitative Evaluation of Approximate Frequent Pattern. (Las Vegas 2008).

5 Astin, M., Griffin, T., Neal, R.D, Rose, P., Hamilton, W. The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. (UK 2011), British Journal of General Practice.

6 Watanabe, T., Nakaya, N., Kurashima, K., Kuriyama, S., Tsubono, Y., Tsuji, I. Constipation, laxative use and risk of colorectal cancer: The Miyagi Cohort Study. *European journal of cancer*, 40, 14 (september 2004), 2109-2115.

7 Kop, R., Hoogendoorn, M., Moons, L.M.G, Numans, M.E., Teije, A.. On the advantages of using dedicated data mining techniques to predict colorectal cancer. *Artificial intelligence in medicine*, 9105 (2015), 133-142.

8 Batal, I., Valizadegan, H., Cooper, G.F., Hauskrecht, M. A temporal pattern mining approach for classifying electronic health record data. *ACM Transactions on Intelligent Systems and Technology*, 5 (August 2012), 3-10.

9 Lewis, R J. An Introduction to Classification and Regression Tree (CART) Analysis. (San Francisco 2000).

10 Svetnik, Vladimir, Liaw, Andy, Tong, Christopher, Culberson, J. Christopher, Sheridan, Robert P., Feuston, Bradley P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of chemical information and modeling*, 43, 6 (november 2003), 1947-1958.

11 Hosmer, D.W., Lemeshow, S. *Applied logistic regression*. John Wiley & Sons, Canada, 2000.

12 Thorsten, J. support vector machines. In *Learning to classify text using support vector machines*. Springer Science+ Business Media, New York, 2002.

13 Hanley, A., McNeil, J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 1 (April 1982), 29-36.

Appendix

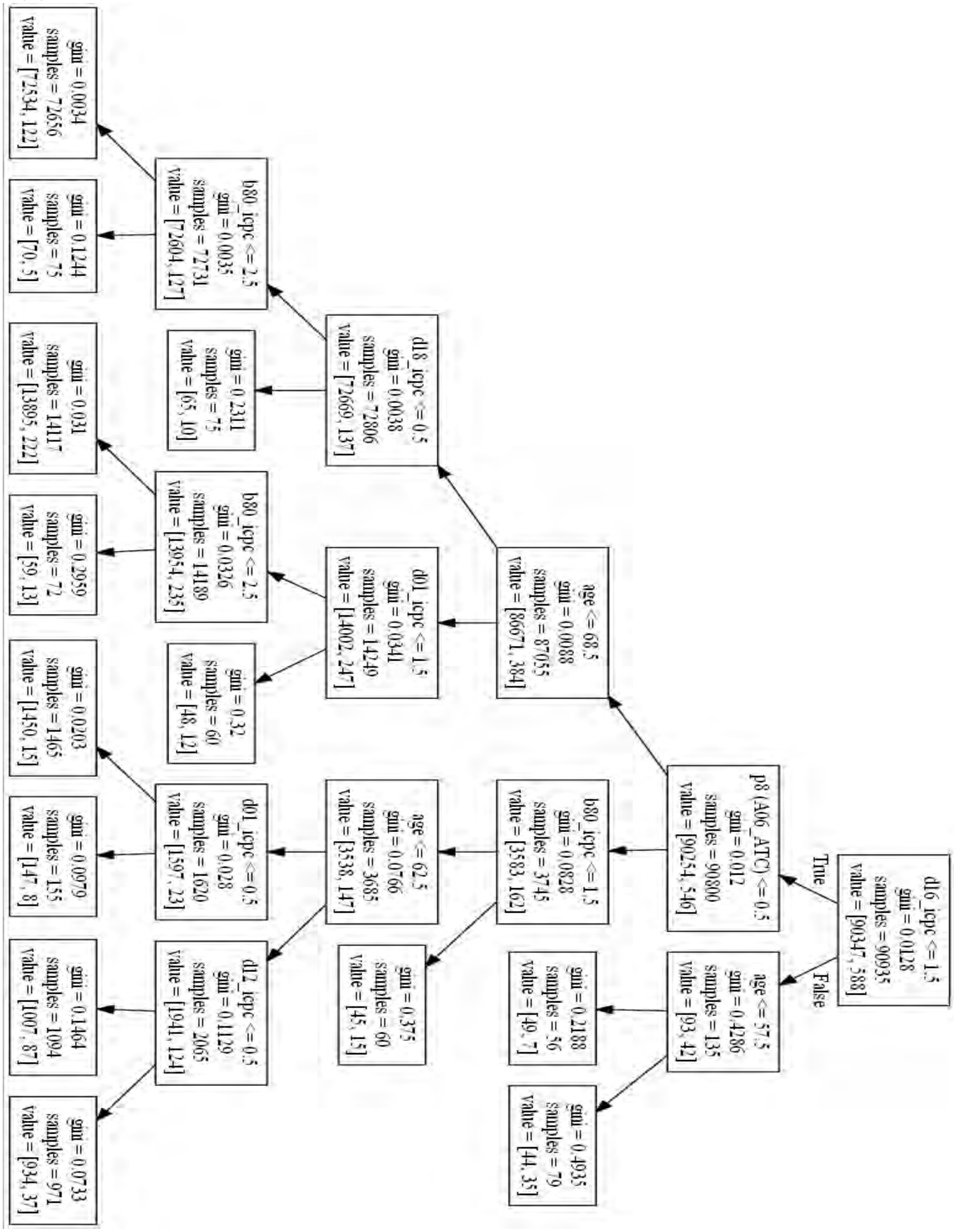


FIGURE 2 : Decision tree for CART with 0.10 support excluding pattern completion times

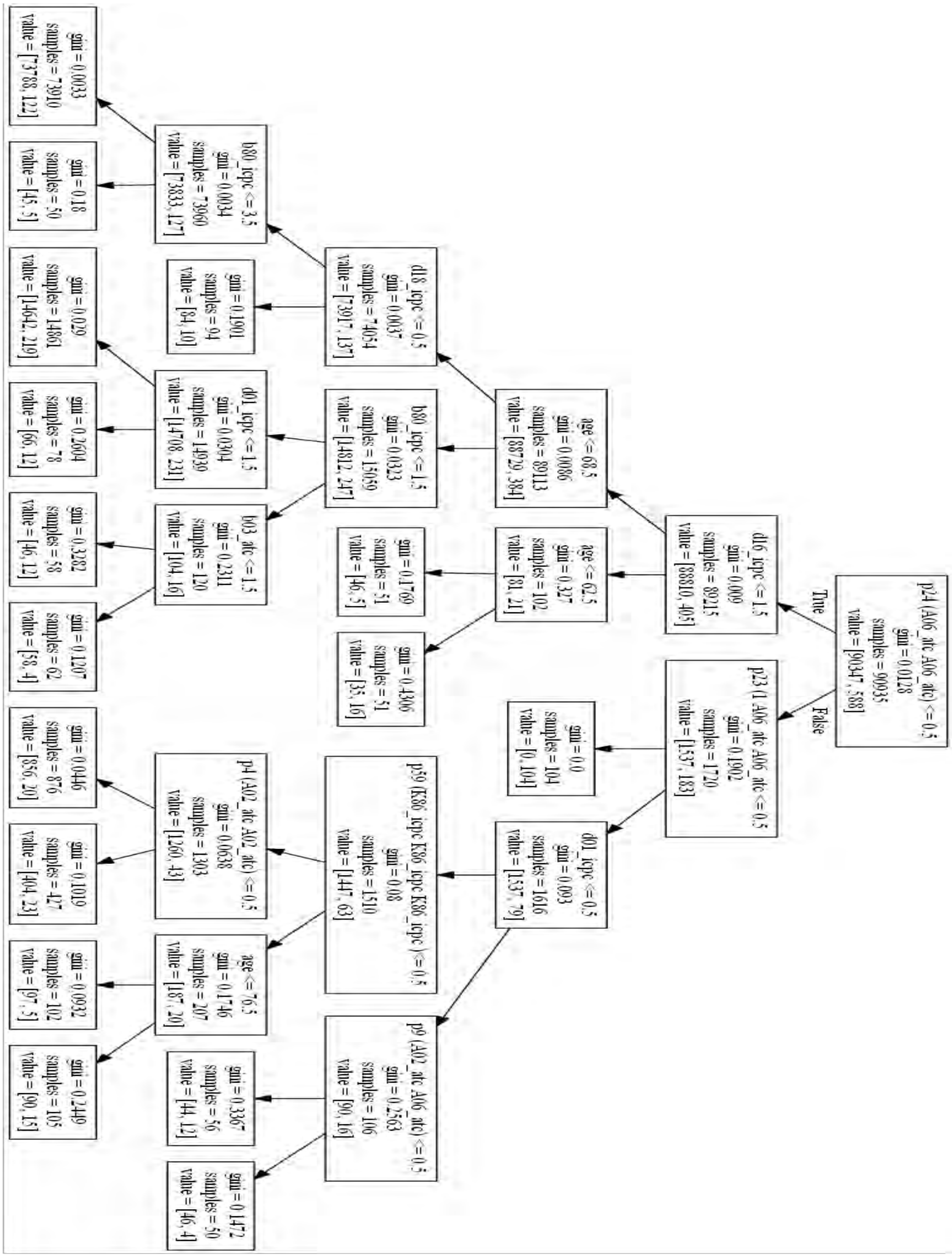


FIGURE 3: Decision tree for CART with 0.10 support including pattern completion times

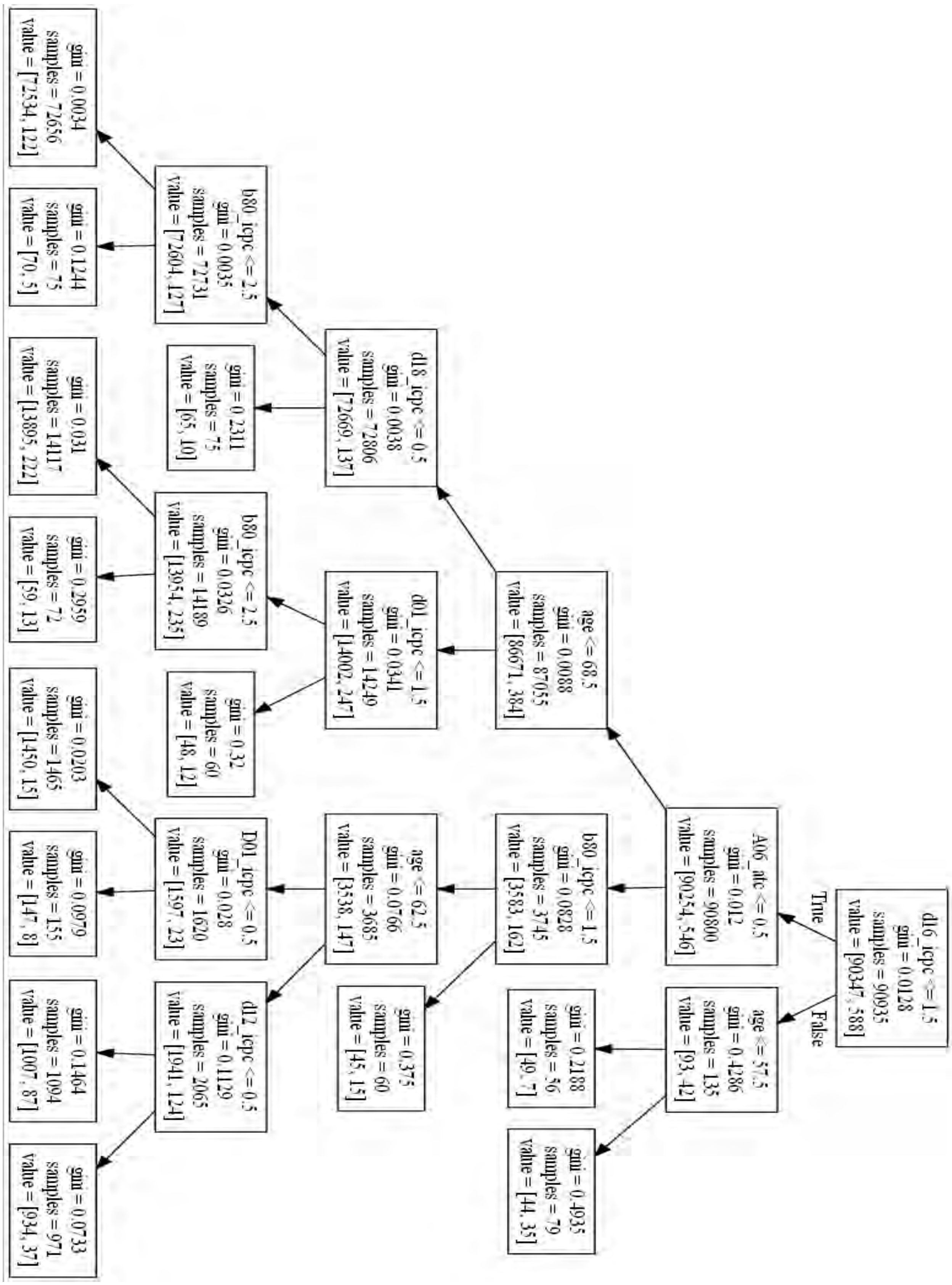


FIGURE 4: Decision tree for CART with 0.05 support excluding pattern completion times

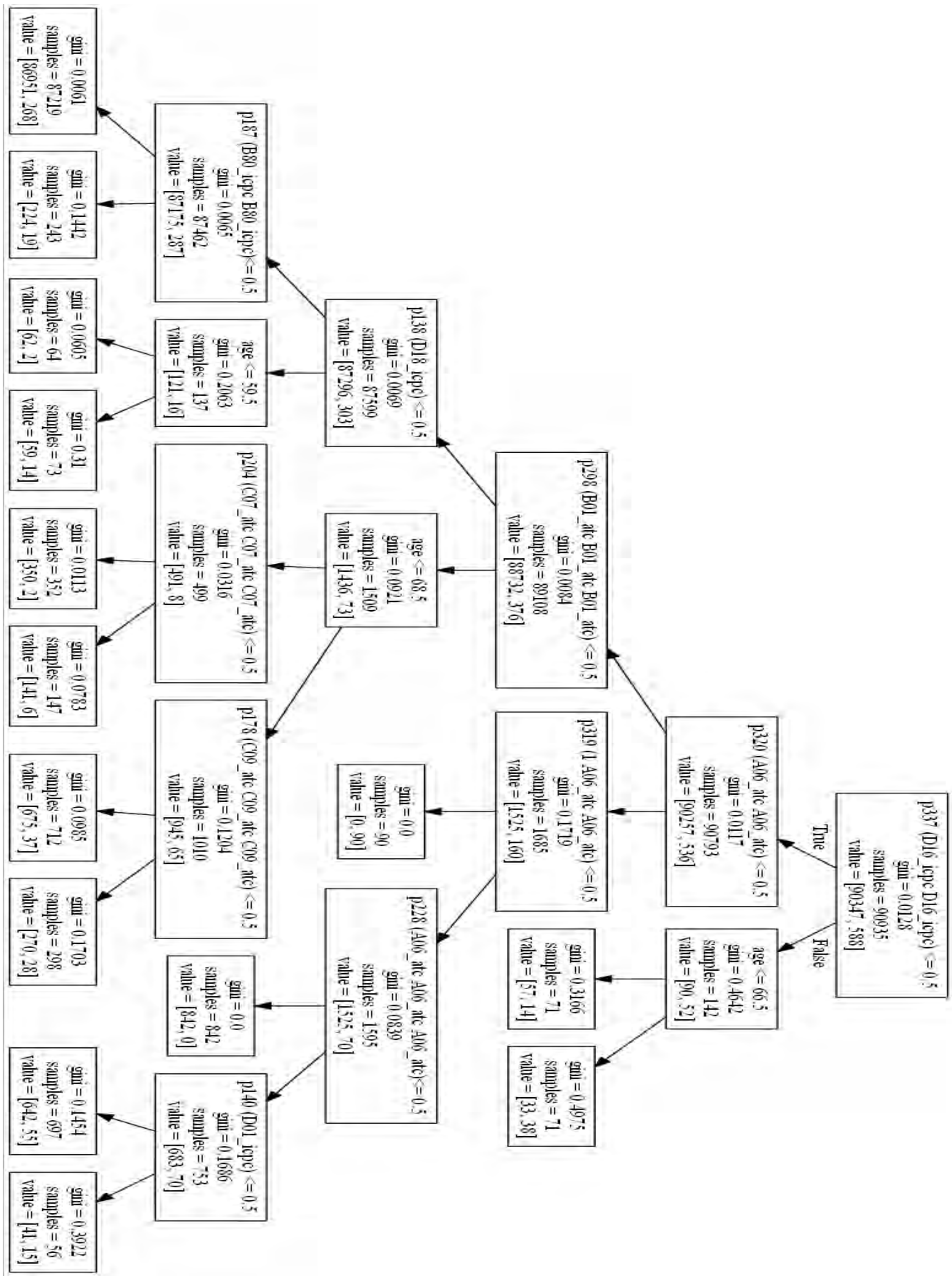


FIGURE 5: Decision tree for CART with 0.05 support including pattern completion times

1-patterns		2-patterns			
p	code	p	code	relation	code
0	K86_icpc	2	A02_atc	b	A02_atc
1	C03_atc	3	C03_atc	c	K86_icpc
5	C09_atc	4	C09_atc	b	C10_atc
6	A02_atc	7	A02_atc	b	A06_atc
8	A06_atc	9	C09_atc	b	K86_icpc
14	R03_atc	10	A99_icpc	b	A99_icpc
15	R44_icpc	11	C09_atc	c	K86_icpc
16	A99_icpc	12	C03_atc	b	C03_atc
20	T90_icpc	13	B01_atc	b	C10_atc
21	C10_atc	17	A06_atc	b	A06_atc
22	A10_atc	18	A10_atc	b	T90_icpc
24	J01_atc	19	C10_atc	b	B01_atc
25	C08_atc	23	C07_atc	b	C07_atc
26	C07_atc	27	K86_icpc	b	K86_icpc
28	N05_atc	29	C09_atc	b	C09_atc
31	B03_atc	30	B01_atc	b	C09_atc
32	N02_atc	33	N05_atc	b	N05_atc
40	B01_atc	34	C10_atc	b	C10_atc
41	D12_icpc	35	B01_atc	b	B01_atc
		36	C10_atc	b	C09_atc
		38	T90_icpc	b	T90_icpc
		39	C03_atc	b	K86_icpc
		42	C09_atc	b	B01_atc

TABLE 15: patterns found by the temporal pattern mining algorithm with a minimum support of 0.10 excluding pattern completion times

1-patterns		2-patterns					3-patterns						
p	code	interval	p	code	relation	code	interval	p	code	relation	code	relation	code
0	K86_icpc	9-16	2	A02_atc	b	A02_atc	9-16	57	K86_icpc	b	K86_icpc	b	K86_icpc
1	C03_atc	0-1	3	A02_atc	b	A02_atc	0-1	58	K86_icpc	b	K86_icpc	b	K86_icpc
7	C09_atc		4	A02_atc	b	A02_atc	0-1	67	T90_icpc	b	T90_icpc	b	T90_icpc
8	A02_atc		5	C03_atc	c	K86_icpc		68	T90_icpc	b	T90_icpc	b	T90_icpc
10	A06_atc		6	C09_atc	b	C10_atc	9-16	69	A99_icpc	b	A99_icpc	b	A99_icpc
20	R03_atc		9	A02_atc	b	A06_atc	0-1	70	A99_icpc	b	A99_icpc	b	A99_icpc
21	R44_icpc		11	C09_atc	b	K86_icpc							
22	A99_icpc	9-16	12	A99_icpc	b	A99_icpc							
27	T90_icpc	0-1	13	A99_icpc	b	A99_icpc							
28	C10_atc		14	A99_icpc	b	A99_icpc							
29	A10_atc		15	C09_atc	c	K86_icpc							
33	J01_atc	9-16	16	C03_atc	b	C03_atc							
34	C08_atc	0-1	17	C03_atc	b	C03_atc							
35	C07_atc		18	C03_atc	b	C03_atc							
39	N05_atc		19	B01_atc	b	C10_atc							
44	B03_atc	0-1	23	A06_atc	b	A06_atc							
45	N02_atc		24	A06_atc	b	A06_atc							
64	B01_atc		25	A10_atc	b	T90_icpc							
65	D12_icpc		26	C10_atc	b	B01_atc							
		9-16	30	C07_atc	b	C07_atc							
		0-1	31	C07_atc	b	C07_atc							
			32	C07_atc	b	C07_atc							
		9-16	36	K86_icpc	b	K86_icpc							
		0-1	37	K86_icpc	b	K86_icpc							
			38	K86_icpc	b	K86_icpc							
		9-16	40	C09_atc	b	C09_atc							
		0-1	41	C09_atc	b	C09_atc							
			42	C09_atc	b	C09_atc							
			43	B01_atc	b	C09_atc							
		9-16	46	N05_atc	b	N05_atc							
		0-1	47	N05_atc	b	N05_atc							
			48	N05_atc	b	N05_atc							
		9-16	49	C10_atc	b	C10_atc							
		0-1	50	C10_atc	b	C10_atc							
			51	C10_atc	b	C10_atc							
		9-16	52	B01_atc	b	B01_atc							
		0-1	53	B01_atc	b	B01_atc							
			54	B01_atc	b	B01_atc							
		0-1	55	C10_atc	b	C09_atc							

			56	C10_atc	b	C09_atc						
		0-1	60	T90_icpc	b	T90_icpc						
			61	T90_icpc	b	T90_icpc						
		0-1	62	C03_atc	b	K86_icpc						
			63	C03_atc	b	K86_icpc						
			66	C09_atc	b	B01_atc						

TABLE 16: patterns found by the temporal pattern mining algorithm with a minimum support of 0.10 including pattern completion times

1-patterns		2-patterns				3-patterns					
p	code	p	code	relation	code	p	code	relation	code	relation	code
0	D02_atc	1	N02_atc	b	N02_atc	15	A10_atc	b	T90_icpc	c	T90_icpc
3	G04_atc	2	T90_icpc	c	T90_icpc	18	C03_atc	c	C03_atc	b	C09_atc
4	H02_atc	5	C03_atc	b	B01_atc	21	N05_atc	b	N05_atc	b	A02_atc
7	C10_atc	6	N05_atc	b	A02_atc	25	A02_atc	b	A02_atc	b	A02_atc
13	K86_icpc	8	C09_atc	b	K86_icpc	27	C09_atc	c	C09_atc	b	K86_icpc
17	J01_atc	9	A10_atc	b	C09_atc	30	B01_atc	b	C10_atc	c	B01_atc
20	T90_icpc	10	C09_atc	b	C07_atc	31	C09_atc	c	C09_atc	b	C10_atc
26	C01_atc	11	R03_atc	c	R95_icpc	32	B01_atc	b	C09_atc	c	C09_atc
28	C07_atc	12	A02_atc	c	B01_atc	35	C09_atc	b	C09_atc	b	B01_atc
33	C08_atc	14	C10_atc	b	A99_icpc	38	C09_atc	b	B01_atc	b	B01_atc
55	B03_atc	16	C09_atc	b	R44_icpc	41	K86_icpc	b	K86_icpc	b	K86_icpc
57	N06_atc	19	C01_atc	b	C01_atc	43	B01_atc	c	B01_atc	b	C07_atc
60	R01_atc	22	C10_atc	b	C09_atc	44	B01_atc	b	C07_atc	c	B01_atc
70	R44_icpc	23	C07_atc	b	A99_icpc	45	B01_atc	b	C07_atc	c	C07_atc
71	D18_icpc	24	C07_atc	b	K86_icpc	61	C07_atc	c	C07_atc	b	K86_icpc
73	D01_icpc	29	C07_atc	b	A02_atc	62	C09_atc	b	K86_icpc	c	K86_icpc
79	U71_icpc	34	C03_atc	b	A99_icpc	63	B01_atc	b	C09_atc	b	B01_atc
82	D12_icpc	36	A06_atc	b	D12_icpc	64	C09_atc	c	A10_atc	b	T90_icpc
91	M01_atc	37	C09_atc	b	A06_atc	68	C07_atc	b	K86_icpc	c	K86_icpc
102	D16_icpc	39	A99_icpc	b	R44_icpc	69	A10_atc	c	A10_atc	b	T90_icpc
104	D06_icpc	40	N05_atc	b	C09_atc	80	B01_atc	b	C10_atc	b	C10_atc
111	A97_icpc	42	B01_atc	c	C10_atc	86	C09_atc	c	C03_atc	b	C09_atc
121	P06_icpc	46	B01_atc	b	C07_atc	87	B01_atc	b	C10_atc	c	C10_atc
122	B01_atc	47	C10_atc	b	C10_atc	92	C09_atc	b	C09_atc	b	C09_atc
124	C09_atc	48	C10_atc	b	T90_icpc	94	C09_atc	b	C09_atc	b	A02_atc
131	R03_atc	49	A02_atc	b	N05_atc	95	A02_atc	b	A99_icpc	b	A99_icpc
135	C03_atc	50	C03_atc	c	C09_atc	100	A10_atc	b	T90_icpc	b	T90_icpc
143	B80_icpc	51	A02_atc	b	A99_icpc	106	C07_atc	b	C07_atc	b	C07_atc
150	A04_icpc	52	K86_icpc	c	K86_icpc	108	B01_atc	b	A02_atc	b	B01_atc
167	S01_atc	53	J01_atc	b	J01_atc	116	C10_atc	c	B01_atc	b	C10_atc
184	R95_icpc	54	C09_atc	b	C03_atc	118	B01_atc	b	C09_atc	b	C09_atc

185	N02_atc	56	C09_atc	b	T90_icpc	119	A06_atc	b	A06_atc	b	A06_atc
187	A10_atc	58	B01_atc	b	A99_icpc	123	N05_atc	b	A02_atc	b	N05_atc
200	A02_atc	59	A02_atc	b	C07_atc	126	C09_atc	b	C10_atc	b	C09_atc
204	T93_icpc	65	C10_atc	b	T93_icpc	127	N05_atc	b	N05_atc	b	N05_atc
209	A03_atc	66	A06_atc	b	A99_icpc	130	B01_atc	c	B01_atc	b	K86_icpc
218	N05_atc	67	B01_atc	b	B01_atc	133	N05_atc	c	N05_atc	b	P06_icpc
219	A99_icpc	72	A02_atc	b	D12_icpc	138	B01_atc	c	A02_atc	b	B01_atc
226	A06_atc	74	D01_icpc	b	D01_icpc	139	C10_atc	c	C09_atc	b	C10_atc
230	D07_atc	75	N05_atc	c	P06_icpc	141	B01_atc	c	B01_atc	b	C09_atc
231	D02_icpc	76	B01_atc	b	R44_icpc	151	C03_atc	b	C09_atc	c	C09_atc
		77	T90_icpc	b	T90_icpc	152	B01_atc	b	B01_atc	b	B01_atc
		78	D12_icpc	b	D12_icpc	154	C09_atc	b	C10_atc	c	C10_atc
		81	C03_atc	b	N05_atc	157	A10_atc	b	A10_atc	b	T90_icpc
		83	A02_atc	b	A06_atc	159	C03_atc	b	K86_icpc	c	K86_icpc
		84	C07_atc	c	K86_icpc	165	B01_atc	b	C10_atc	b	B01_atc
		85	A02_atc	b	A02_atc	170	C09_atc	b	C10_atc	c	C09_atc
		88	C10_atc	b	A06_atc	175	C09_atc	b	T90_icpc	b	T90_icpc
		89	B01_atc	b	C01_atc	183	K86_icpc	b	A99_icpc	b	K86_icpc
		90	C03_atc	b	K86_icpc	186	B01_atc	b	B01_atc	b	C10_atc
		93	B01_atc	b	T90_icpc	188	B01_atc	b	B01_atc	b	C09_atc
		96	A10_atc	b	A10_atc	192	C03_atc	b	C09_atc	c	C03_atc
		97	B80_icpc	b	B80_icpc	193	B01_atc	b	B01_atc	b	A02_atc
		98	B01_atc	c	K86_icpc	194	B01_atc	b	C09_atc	c	B01_atc
		99	R03_atc	b	R95_icpc	197	A99_icpc	b	A99_icpc	b	A99_icpc
		101	A06_atc	b	C09_atc	198	C07_atc	c	B01_atc	b	C07_atc
		103	A10_atc	b	A02_atc	202	C03_atc	c	C03_atc	b	K86_icpc
		105	C07_atc	b	C07_atc	205	C10_atc	b	B01_atc	b	C10_atc
		107	K86_icpc	b	R44_icpc	212	B01_atc	c	B01_atc	b	C10_atc
		109	C09_atc	b	A10_atc	214	C09_atc	b	B01_atc	b	C09_atc
		110	C09_atc	b	A99_icpc	222	C09_atc	c	B01_atc	b	C09_atc
		112	N05_atc	b	K86_icpc	223	T90_icpc	b	T90_icpc	b	T90_icpc
		113	A02_atc	c	C09_atc						
		114	A02_atc	b	B01_atc						
		115	A06_atc	c	D12_icpc						
		117	R44_icpc	b	K86_icpc						
		120	C08_atc	b	C08_atc						
		125	C03_atc	c	K86_icpc						
		128	C09_atc	c	T90_icpc						
		129	C03_atc	b	C03_atc						
		132	N05_atc	b	N05_atc						
		134	C09_atc	b	A02_atc						

		136	N05_atc	b	P06_icpc						
		137	C09_atc	b	C10_atc						
		140	C10_atc	c	K86_icpc						
		142	B01_atc	c	C07_atc						
		144	A99_icpc	b	A99_icpc						
		145	C03_atc	b	C09_atc						
		146	C10_atc	c	T90_icpc						
		147	B03_atc	b	B03_atc						
		148	B01_atc	b	C09_atc						
		149	N05_atc	b	A99_icpc						
		153	C09_atc	c	C10_atc						
		155	B01_atc	b	N05_atc						
		156	C07_atc	b	C03_atc						
		158	C09_atc	b	B01_atc						
		160	A02_atc	b	T90_icpc						
		161	A06_atc	b	A06_atc						
		162	A06_atc	b	A02_atc						
		163	B01_atc	b	A06_atc						
		164	C03_atc	b	A02_atc						
		166	K86_icpc	b	K86_icpc						
		168	C09_atc	c	K86_icpc						
		169	D16_icpc	b	D16_icpc						
		171	C07_atc	b	C10_atc						
		172	N06_atc	b	N06_atc						
		173	C09_atc	b	C09_atc						
		174	C09_atc	b	N05_atc						
		176	N05_atc	b	B01_atc						
		177	C10_atc	b	A02_atc						
		178	C10_atc	b	B01_atc						
		179	A99_icpc	b	K86_icpc						
		180	C07_atc	b	C09_atc						
		181	B01_atc	b	A02_atc						
		182	A02_atc	b	N02_atc						
		189	A02_atc	c	N05_atc						
		190	A10_atc	b	T90_icpc						
		191	A02_atc	c	C03_atc						
		195	R44_icpc	b	A99_icpc						
		196	C03_atc	b	A06_atc						
		199	B01_atc	b	C10_atc						
		201	C10_atc	b	C07_atc						
		203	C07_atc	b	B01_atc						

		206	A02_atc	b	C10_atc						
		207	B01_atc	b	K86_icpc						
		208	R03_atc	b	R03_atc						
		210	C03_atc	b	C07_atc						
		211	B01_atc	c	C09_atc						
		213	A02_atc	b	C03_atc						
		215	C03_atc	c	C07_atc						
		216	K86_icpc	b	A99_icpc						
		217	C10_atc	c	T93_icpc						
		220	R95_icpc	b	R95_icpc						
		221	A10_atc	c	T90_icpc						
		224	A02_atc	b	K86_icpc						
		225	C10_atc	b	K86_icpc						
		227	B01_atc	c	C03_atc						
		228	A02_atc	b	C09_atc						
		229	B01_atc	b	C03_atc						

TABLE 17: patterns found by the temporal pattern mining algorithm with a minimum support of 0.05 excluding pattern completion times

1-patterns		2-patterns					3-patterns					
p	code	interval	p	code	relation	code	interval	p	code	relation	code	relation	code
0	D02_atc	0-1	1	N02_atc	b	N02_atc	9-16	22	A10_atc	b	T90_icpc	c	T90_icpc
4	G04_atc		2	N02_atc	b	N02_atc	0-1	23	A10_atc	b	T90_icpc	c	T90_icpc
5	H02_atc		3	T90_icpc	c	T90_icpc		24	A10_atc	b	T90_icpc	c	T90_icpc
10	C10_atc	9-16	6	C03_atc	b	B01_atc	0-1	27	C03_atc	c	C03_atc	b	C09_atc
20	K86_icpc	0-1	7	C03_atc	b	B01_atc		28	C03_atc	c	C03_atc	b	C09_atc
26	J01_atc		8	C03_atc	b	B01_atc		32	N05_atc	b	N05_atc	b	A02_atc
31	T90_icpc		9	N05_atc	b	A02_atc	9-16	40	A02_atc	b	A02_atc	b	A02_atc
44	C01_atc	9-16	11	C09_atc	b	K86_icpc	0-1	41	A02_atc	b	A02_atc	b	A02_atc
48	C07_atc	0-1	12	C09_atc	b	K86_icpc	5-8	42	A02_atc	b	A02_atc	b	A02_atc
59	C08_atc		13	C09_atc	b	K86_icpc		43	A02_atc	b	A02_atc	b	A02_atc
106	B03_atc	0-1	14	A10_atc	b	C09_atc	9-16	45	C09_atc	c	C09_atc	b	K86_icpc
110	N06_atc		15	A10_atc	b	C09_atc	0-1	46	C09_atc	c	C09_atc	b	K86_icpc
114	R01_atc	0-1	16	C09_atc	b	C07_atc		47	C09_atc	c	C09_atc	b	K86_icpc
137	R44_icpc		17	C09_atc	b	C07_atc	9-16	50	B01_atc	b	C10_atc	c	B01_atc
138	D18_icpc		18	R03_atc	c	R95_icpc	0-1	51	B01_atc	b	C10_atc	c	B01_atc
140	D01_icpc		19	A02_atc	c	B01_atc		52	B01_atc	b	C10_atc	c	B01_atc
151	U71_icpc		21	C10_atc	b	A99_icpc	9-16	53	C09_atc	c	C09_atc	b	C10_atc
156	D12_icpc		25	C09_atc	b	R44_icpc	0-1	54	C09_atc	c	C09_atc	b	C10_atc
174	M01_atc	0-1	29	C01_atc	b	C01_atc		55	C09_atc	c	C09_atc	b	C10_atc
196	D16_icpc		30	C01_atc	b	C01_atc	9-16	56	B01_atc	b	C09_atc	c	C09_atc

198	D06_icpc	9-16	33	C10_atc	b	C09_atc	0-1	57	B01_atc	b	C09_atc	c	C09_atc
212	A97_icpc	0-1	34	C10_atc	b	C09_atc		58	B01_atc	b	C09_atc	c	C09_atc
232	P06_icpc		35	C10_atc	b	C09_atc	9-16	62	C09_atc	b	C09_atc	b	B01_atc
233	B01_atc		36	C07_atc	b	A99_icpc	0-1	63	C09_atc	b	C09_atc	b	B01_atc
237	C09_atc	9-16	37	C07_atc	b	K86_icpc		64	C09_atc	b	C09_atc	b	B01_atc
252	R03_atc	0-1	38	C07_atc	b	K86_icpc	9-16	68	C09_atc	b	B01_atc	b	B01_atc
261	C03_atc		39	C07_atc	b	K86_icpc	0-1	69	C09_atc	b	B01_atc	b	B01_atc
277	B80_icpc		49	C07_atc	b	A02_atc		70	C09_atc	b	B01_atc	b	B01_atc
293	A04_icpc	0-1	60	C03_atc	b	A99_icpc	9-16	73	K86_icpc	b	K86_icpc	b	K86_icpc
334	S01_atc		61	C03_atc	b	A99_icpc	17-32	74	K86_icpc	b	K86_icpc	b	K86_icpc
371	R95_icpc	0-1	65	A06_atc	b	D12_icpc	0-1	75	K86_icpc	b	K86_icpc	b	K86_icpc
372	N02_atc		66	A06_atc	b	D12_icpc	5-8	76	K86_icpc	b	K86_icpc	b	K86_icpc
376	A10_atc		67	C09_atc	b	A06_atc		77	K86_icpc	b	K86_icpc	b	K86_icpc
406	A02_atc		71	A99_icpc	b	R44_icpc	9-16	79	B01_atc	c	B01_atc	b	C07_atc
415	T93_icpc		72	N05_atc	b	C09_atc	0-1	80	B01_atc	c	B01_atc	b	C07_atc
427	A03_atc		78	B01_atc	c	C10_atc		81	B01_atc	c	B01_atc	b	C07_atc
442	N05_atc	9-16	88	B01_atc	b	C07_atc	9-16	82	B01_atc	b	C07_atc	c	B01_atc
443	A99_icpc	0-1	89	B01_atc	b	C07_atc	0-1	83	B01_atc	b	C07_atc	c	B01_atc
458	A06_atc		90	B01_atc	b	C07_atc		84	B01_atc	b	C07_atc	c	B01_atc
466	D07_atc	9-16	91	C10_atc	b	C10_atc	9-16	85	B01_atc	b	C07_atc	c	C07_atc
467	D02_icpc	0-1	92	C10_atc	b	C10_atc	0-1	86	B01_atc	b	C07_atc	c	C07_atc
			93	C10_atc	b	C10_atc		87	B01_atc	b	C07_atc	c	C07_atc
		0-1	94	C10_atc	b	T90_icpc	9-16	115	C07_atc	c	C07_atc	b	K86_icpc
			95	C10_atc	b	T90_icpc	0-1	116	C07_atc	c	C07_atc	b	K86_icpc
		0-1	96	A02_atc	b	N05_atc		117	C07_atc	c	C07_atc	b	K86_icpc
			97	A02_atc	b	N05_atc	9-16	118	C09_atc	b	K86_icpc	c	K86_icpc
			98	C03_atc	c	C09_atc	0-1	119	C09_atc	b	K86_icpc	c	K86_icpc
		0-1	99	A02_atc	b	A99_icpc		120	C09_atc	b	K86_icpc	c	K86_icpc
			100	A02_atc	b	A99_icpc	9-16	121	B01_atc	b	C09_atc	b	B01_atc
			101	K86_icpc	c	K86_icpc	0-1	122	B01_atc	b	C09_atc	b	B01_atc
		0-1	102	J01_atc	b	J01_atc		123	B01_atc	b	C09_atc	b	B01_atc
			103	J01_atc	b	J01_atc		124	C09_atc	c	A10_atc	b	T90_icpc
		0-1	104	C09_atc	b	C03_atc	9-16	131	C07_atc	b	K86_icpc	c	K86_icpc
			105	C09_atc	b	C03_atc	0-1	132	C07_atc	b	K86_icpc	c	K86_icpc
		9-16	107	C09_atc	b	T90_icpc		133	C07_atc	b	K86_icpc	c	K86_icpc
		0-1	108	C09_atc	b	T90_icpc	9-16	134	A10_atc	c	A10_atc	b	T90_icpc
			109	C09_atc	b	T90_icpc	0-1	135	A10_atc	c	A10_atc	b	T90_icpc
		0-1	111	B01_atc	b	A99_icpc		136	A10_atc	c	A10_atc	b	T90_icpc
			112	B01_atc	b	A99_icpc	9-16	152	B01_atc	b	C10_atc	b	C10_atc
			113	A02_atc	b	C07_atc	0-1	153	B01_atc	b	C10_atc	b	C10_atc
			125	C10_atc	b	T93_icpc		154	B01_atc	b	C10_atc	b	C10_atc

		0-1	126	A06_atc	b	A99_icpc	0-1	164	C09_atc	c	C03_atc	b	C09_atc
			127	A06_atc	b	A99_icpc		165	C09_atc	c	C03_atc	b	C09_atc
		9-16	128	B01_atc	b	B01_atc	9-16	166	B01_atc	b	C10_atc	c	C10_atc
		0-1	129	B01_atc	b	B01_atc	0-1	167	B01_atc	b	C10_atc	c	C10_atc
			130	B01_atc	b	B01_atc		168	B01_atc	b	C10_atc	c	C10_atc
			139	A02_atc	b	D12_icpc	9-16	175	C09_atc	b	C09_atc	b	C09_atc
		0-1	141	D01_icpc	b	D01_icpc	17-32	176	C09_atc	b	C09_atc	b	C09_atc
			142	D01_icpc	b	D01_icpc	0-1	177	C09_atc	b	C09_atc	b	C09_atc
			143	N05_atc	c	P06_icpc		178	C09_atc	b	C09_atc	b	C09_atc
			144	B01_atc	b	R44_icpc		180	C09_atc	b	C09_atc	b	A02_atc
		9-16	145	T90_icpc	b	T90_icpc	0-1	181	A02_atc	b	A99_icpc	b	A99_icpc
		17-32	146	T90_icpc	b	T90_icpc		182	A02_atc	b	A99_icpc	b	A99_icpc
		0-1	147	T90_icpc	b	T90_icpc	9-16	191	A10_atc	b	T90_icpc	b	T90_icpc
			148	T90_icpc	b	T90_icpc	0-1	192	A10_atc	b	T90_icpc	b	T90_icpc
		0-1	149	D12_icpc	b	D12_icpc		193	A10_atc	b	T90_icpc	b	T90_icpc
			150	D12_icpc	b	D12_icpc	9-16	202	C07_atc	b	C07_atc	b	C07_atc
			155	C03_atc	b	N05_atc	0-1	203	C07_atc	b	C07_atc	b	C07_atc
		0-1	157	A02_atc	b	A06_atc		204	C07_atc	b	C07_atc	b	C07_atc
			158	A02_atc	b	A06_atc	9-16	206	B01_atc	b	A02_atc	b	B01_atc
			159	C07_atc	c	K86_icpc	0-1	207	B01_atc	b	A02_atc	b	B01_atc
		9-16	160	A02_atc	b	A02_atc		208	B01_atc	b	A02_atc	b	B01_atc
		0-1	161	A02_atc	b	A02_atc	9-16	218	C10_atc	c	B01_atc	b	C10_atc
		5-8	162	A02_atc	b	A02_atc	0-1	219	C10_atc	c	B01_atc	b	C10_atc
			163	A02_atc	b	A02_atc		220	C10_atc	c	B01_atc	b	C10_atc
			169	C10_atc	b	A06_atc	9-16	223	B01_atc	b	C09_atc	b	C09_atc
			170	B01_atc	b	C01_atc	0-1	224	B01_atc	b	C09_atc	b	C09_atc
		9-16	171	C03_atc	b	K86_icpc		225	B01_atc	b	C09_atc	b	C09_atc
		0-1	172	C03_atc	b	K86_icpc	9-16	226	A06_atc	b	A06_atc	b	A06_atc
			173	C03_atc	b	K86_icpc	0-1	227	A06_atc	b	A06_atc	b	A06_atc
			179	B01_atc	b	T90_icpc		228	A06_atc	b	A06_atc	b	A06_atc
		9-16	183	A10_atc	b	A10_atc	9-16	234	N05_atc	b	A02_atc	b	N05_atc
		0-1	184	A10_atc	b	A10_atc	0-1	235	N05_atc	b	A02_atc	b	N05_atc
			185	A10_atc	b	A10_atc		236	N05_atc	b	A02_atc	b	N05_atc
		0-1	186	B80_icpc	b	B80_icpc	9-16	239	C09_atc	b	C10_atc	b	C09_atc
			187	B80_icpc	b	B80_icpc	0-1	240	C09_atc	b	C10_atc	b	C09_atc
			188	B01_atc	c	K86_icpc		241	C09_atc	b	C10_atc	b	C09_atc
		0-1	189	R03_atc	b	R95_icpc	9-16	242	N05_atc	b	N05_atc	b	N05_atc
			190	R03_atc	b	R95_icpc	17-32	243	N05_atc	b	N05_atc	b	N05_atc
		0-1	194	A06_atc	b	C09_atc	0-1	244	N05_atc	b	N05_atc	b	N05_atc
			195	A06_atc	b	C09_atc	5-8	245	N05_atc	b	N05_atc	b	N05_atc
			197	A10_atc	b	A02_atc		246	N05_atc	b	N05_atc	b	N05_atc

		9-16	199	C07_atc	b	C07_atc		251	B01_atc	c	B01_atc	b	K86_icpc
		0-1	200	C07_atc	b	C07_atc		258	N05_atc	c	N05_atc	b	P06_icpc
			201	C07_atc	b	C07_atc	0-1	267	B01_atc	c	A02_atc	b	B01_atc
			205	K86_icpc	b	R44_icpc		268	B01_atc	c	A02_atc	b	B01_atc
		0-1	209	C09_atc	b	A10_atc	9-16	269	C10_atc	c	C09_atc	b	C10_atc
			210	C09_atc	b	A10_atc	0-1	270	C10_atc	c	C09_atc	b	C10_atc
			211	C09_atc	b	A99_icpc		271	C10_atc	c	C09_atc	b	C10_atc
			213	N05_atc	b	K86_icpc	9-16	273	B01_atc	c	B01_atc	b	C09_atc
			214	A02_atc	c	C09_atc	0-1	274	B01_atc	c	B01_atc	b	C09_atc
		0-1	215	A02_atc	b	B01_atc		275	B01_atc	c	B01_atc	b	C09_atc
			216	A02_atc	b	B01_atc	0-1	294	C03_atc	b	C09_atc	c	C09_atc
			217	A06_atc	c	D12_icpc		295	C03_atc	b	C09_atc	c	C09_atc
		0-1	221	R44_icpc	b	K86_icpc	9-16	296	B01_atc	b	B01_atc	b	B01_atc
			222	R44_icpc	b	K86_icpc	0-1	297	B01_atc	b	B01_atc	b	B01_atc
		9-16	229	C08_atc	b	C08_atc		298	B01_atc	b	B01_atc	b	B01_atc
		0-1	230	C08_atc	b	C08_atc	9-16	300	C09_atc	b	C10_atc	c	C10_atc
			231	C08_atc	b	C08_atc	0-1	301	C09_atc	b	C10_atc	c	C10_atc
			238	C03_atc	c	K86_icpc		302	C09_atc	b	C10_atc	c	C10_atc
			247	C09_atc	c	T90_icpc	9-16	307	A10_atc	b	A10_atc	b	T90_icpc
		9-16	248	C03_atc	b	C03_atc	0-1	308	A10_atc	b	A10_atc	b	T90_icpc
		0-1	249	C03_atc	b	C03_atc		309	A10_atc	b	A10_atc	b	T90_icpc
			250	C03_atc	b	C03_atc	9-16	313	C03_atc	b	K86_icpc	c	K86_icpc
		9-16	253	N05_atc	b	N05_atc	0-1	314	C03_atc	b	K86_icpc	c	K86_icpc
		17-32	254	N05_atc	b	N05_atc		315	C03_atc	b	K86_icpc	c	K86_icpc
		0-1	255	N05_atc	b	N05_atc	9-16	326	B01_atc	b	C10_atc	b	B01_atc
		5-8	256	N05_atc	b	N05_atc	0-1	327	B01_atc	b	C10_atc	b	B01_atc
			257	N05_atc	b	N05_atc		328	B01_atc	b	C10_atc	b	B01_atc
		0-1	259	C09_atc	b	A02_atc	9-16	338	C09_atc	b	C10_atc	c	C09_atc
			260	C09_atc	b	A02_atc	0-1	339	C09_atc	b	C10_atc	c	C09_atc
		0-1	262	N05_atc	b	P06_icpc		340	C09_atc	b	C10_atc	c	C09_atc
			263	N05_atc	b	P06_icpc	0-1	351	C09_atc	b	T90_icpc	b	T90_icpc
		9-16	264	C09_atc	b	C10_atc		352	C09_atc	b	T90_icpc	b	T90_icpc
		0-1	265	C09_atc	b	C10_atc	9-16	368	K86_icpc	b	A99_icpc	b	K86_icpc
			266	C09_atc	b	C10_atc	0-1	369	K86_icpc	b	A99_icpc	b	K86_icpc
			272	C10_atc	c	K86_icpc		370	K86_icpc	b	A99_icpc	b	K86_icpc
			276	B01_atc	c	C07_atc	9-16	373	B01_atc	b	B01_atc	b	C10_atc
		9-16	278	A99_icpc	b	A99_icpc	0-1	374	B01_atc	b	B01_atc	b	C10_atc
		17-32	279	A99_icpc	b	A99_icpc		375	B01_atc	b	B01_atc	b	C10_atc
		0-1	280	A99_icpc	b	A99_icpc	9-16	377	B01_atc	b	B01_atc	b	C09_atc
		5-8	281	A99_icpc	b	A99_icpc	0-1	378	B01_atc	b	B01_atc	b	C09_atc
			282	A99_icpc	b	A99_icpc		379	B01_atc	b	B01_atc	b	C09_atc

		0-1	283	C03_atc	b	C09_atc	0-1	385	C03_atc	b	C09_atc	c	C03_atc
			284	C03_atc	b	C09_atc		386	C03_atc	b	C09_atc	c	C03_atc
			285	C10_atc	c	T90_icpc	9-16	387	B01_atc	b	B01_atc	b	A02_atc
		0-1	286	B03_atc	b	B03_atc		388	B01_atc	b	B01_atc	b	A02_atc
			287	B03_atc	b	B03_atc	9-16	389	B01_atc	b	C09_atc	c	B01_atc
		9-16	288	B01_atc	b	C09_atc	0-1	390	B01_atc	b	C09_atc	c	B01_atc
		0-1	289	B01_atc	b	C09_atc		391	B01_atc	b	C09_atc	c	B01_atc
			290	B01_atc	b	C09_atc	9-16	395	A99_icpc	b	A99_icpc	b	A99_icpc
		0-1	291	N05_atc	b	A99_icpc	17-32	396	A99_icpc	b	A99_icpc	b	A99_icpc
			292	N05_atc	b	A99_icpc	0-1	397	A99_icpc	b	A99_icpc	b	A99_icpc
			299	C09_atc	c	C10_atc	5-8	398	A99_icpc	b	A99_icpc	b	A99_icpc
		0-1	303	B01_atc	b	N05_atc		399	A99_icpc	b	A99_icpc	b	A99_icpc
			304	B01_atc	b	N05_atc	9-16	400	C07_atc	c	B01_atc	b	C07_atc
		0-1	305	C07_atc	b	C03_atc	0-1	401	C07_atc	c	B01_atc	b	C07_atc
			306	C07_atc	b	C03_atc		402	C07_atc	c	B01_atc	b	C07_atc
		9-16	310	C09_atc	b	B01_atc	9-16	409	C03_atc	c	C03_atc	b	K86_icpc
		0-1	311	C09_atc	b	B01_atc	0-1	410	C03_atc	c	C03_atc	b	K86_icpc
			312	C09_atc	b	B01_atc		411	C03_atc	c	C03_atc	b	K86_icpc
		0-1	316	A02_atc	b	T90_icpc	9-16	416	C10_atc	b	B01_atc	b	C10_atc
			317	A02_atc	b	T90_icpc	0-1	417	C10_atc	b	B01_atc	b	C10_atc
		9-16	318	A06_atc	b	A06_atc		418	C10_atc	b	B01_atc	b	C10_atc
		0-1	319	A06_atc	b	A06_atc	9-16	431	B01_atc	c	B01_atc	b	C10_atc
			320	A06_atc	b	A06_atc	0-1	432	B01_atc	c	B01_atc	b	C10_atc
		0-1	321	A06_atc	b	A02_atc		433	B01_atc	c	B01_atc	b	C10_atc
			322	A06_atc	b	A02_atc	9-16	436	C09_atc	b	B01_atc	b	C09_atc
			323	B01_atc	b	A06_atc	0-1	437	C09_atc	b	B01_atc	b	C09_atc
		0-1	324	C03_atc	b	A02_atc		438	C09_atc	b	B01_atc	b	C09_atc
			325	C03_atc	b	A02_atc	9-16	447	C09_atc	c	B01_atc	b	C09_atc
		9-16	329	K86_icpc	b	K86_icpc	0-1	448	C09_atc	c	B01_atc	b	C09_atc
		17-32	330	K86_icpc	b	K86_icpc		449	C09_atc	c	B01_atc	b	C09_atc
		0-1	331	K86_icpc	b	K86_icpc	9-16	450	T90_icpc	b	T90_icpc	b	T90_icpc
		5-8	332	K86_icpc	b	K86_icpc	17-32	451	T90_icpc	b	T90_icpc	b	T90_icpc
			333	K86_icpc	b	K86_icpc	0-1	452	T90_icpc	b	T90_icpc	b	T90_icpc
			335	C09_atc	c	K86_icpc		453	T90_icpc	b	T90_icpc	b	T90_icpc
		0-1	336	D16_icpc	b	D16_icpc							
			337	D16_icpc	b	D16_icpc							
		0-1	341	C07_atc	b	C10_atc							
			342	C07_atc	b	C10_atc							
		9-16	343	N06_atc	b	N06_atc							
		0-1	344	N06_atc	b	N06_atc							
			345	N06_atc	b	N06_atc							

		9-16	346	C09_atc	b	C09_atc							
		17-32	347	C09_atc	b	C09_atc							
		0-1	348	C09_atc	b	C09_atc							
			349	C09_atc	b	C09_atc							
			350	C09_atc	b	N05_atc							
		0-1	353	N05_atc	b	B01_atc							
			354	N05_atc	b	B01_atc							
			355	C10_atc	b	A02_atc							
		9-16	356	C10_atc	b	B01_atc							
		0-1	357	C10_atc	b	B01_atc							
			358	C10_atc	b	B01_atc							
		0-1	359	A99_icpc	b	K86_icpc							
			360	A99_icpc	b	K86_icpc							
		0-1	361	C07_atc	b	C09_atc							
			362	C07_atc	b	C09_atc							
		9-16	363	B01_atc	b	A02_atc							
		0-1	364	B01_atc	b	A02_atc							
			365	B01_atc	b	A02_atc							
		0-1	366	A02_atc	b	N02_atc							
			367	A02_atc	b	N02_atc							
			380	A02_atc	c	N05_atc							
		9-16	381	A10_atc	b	T90_icpc							
		0-1	382	A10_atc	b	T90_icpc							
			383	A10_atc	b	T90_icpc							
			384	A02_atc	c	C03_atc							
		0-1	392	R44_icpc	b	A99_icpc							
			393	R44_icpc	b	A99_icpc							
			394	C03_atc	b	A06_atc							
		9-16	403	B01_atc	b	C10_atc							
		0-1	404	B01_atc	b	C10_atc							
			405	B01_atc	b	C10_atc							
		0-1	407	C10_atc	b	C07_atc							
			408	C10_atc	b	C07_atc							
		9-16	412	C07_atc	b	B01_atc							
		0-1	413	C07_atc	b	B01_atc							
			414	C07_atc	b	B01_atc							
		0-1	419	A02_atc	b	C10_atc							
			420	A02_atc	b	C10_atc							
		9-16	421	B01_atc	b	K86_icpc							
		0-1	422	B01_atc	b	K86_icpc							
			423	B01_atc	b	K86_icpc							

		9-16	424	R03_atc	b	R03_atc							
		0-1	425	R03_atc	b	R03_atc							
			426	R03_atc	b	R03_atc							
		0-1	428	C03_atc	b	C07_atc							
			429	C03_atc	b	C07_atc							
			430	B01_atc	c	C09_atc							
		0-1	434	A02_atc	b	C03_atc							
			435	A02_atc	b	C03_atc							
			439	C03_atc	c	C07_atc							
			440	K86_icpc	b	A99_icpc							
			441	C10_atc	c	T93_icpc							
		0-1	444	R95_icpc	b	R95_icpc							
			445	R95_icpc	b	R95_icpc							
			446	A10_atc	c	T90_icpc							
		0-1	454	A02_atc	b	K86_icpc							
			455	A02_atc	b	K86_icpc							
		0-1	456	C10_atc	b	K86_icpc							
			457	C10_atc	b	K86_icpc							
			459	B01_atc	c	C03_atc							
		9-16	460	A02_atc	b	C09_atc							
		0-1	461	A02_atc	b	C09_atc							
			462	A02_atc	b	C09_atc							
		9-16	463	B01_atc	b	C03_atc							
		0-1	464	B01_atc	b	C03_atc							
			465	B01_atc	b	C03_atc							

TABLE 18: patterns found by the temporal pattern mining algorithm with a minimum support of 0.05 including pattern completion times