Review of Literature on Recursive Partitioning and its applications in various areas

Abhijeet Mankar[1], Sudhakar Bhoite[2]

{abhijeetmankar@gmail.com[1], sdbhoite05@gmail.com[2]}

Tuljaram Chaturchand College, Baramati(Dist. Pune)[1], CSIBER, Kolhapur[2]

Abstract. Prediction is one of the most dominant challenges in data analhysis. Prediction is required in almost every walk of life. Prediction is also some kind of decision and hence it should be possible to use tree structures to represent prediction models. The process of construction of a tree involves partitioning data recursively. There are various application domains of prediction using recursive partitioning like diagnostic analysis, insurance sector, infrastructure, social science, atmospheric study, telecommunications industry.

Keywords: recursive partitioning, CART, decision tree, prediction

1 Introduction

The recursive partitioning approach reflects a tree representation. The prediction is located at the leaf nodes in a tree. Prediction is based on the principle of generalization. Prediction is made under the assumption that the past will repeat itself. Trees designed for classfication predicts categorical dependent variable whereas trees designed for regression predicts continuous dependent variable. Through the years there has been significant research in this field. In this paper, we discuss the various approaches used in recursive partitioning. Also, the features of these approaches and the terms used are discussed. The application areas of recursive partitioning are categorized into different sectors based on its use is discussed in this paper.

2 Overview of recursive partitioning methods

AID(Morgan, Sonquist(1963)) : The authors proposed AID in this paper. It explain about the interaction among variables. The authors highlighted the existence of intercorrelations between many of the explanatory factors to be used in the analysis. The authors further mentioned that how two explanatory factors are correlated with one another, but they do not interact with each other. Also they mentioned that how two explanatory factors interact with one another but are not correlated. The authors stated the possibility of two highly correlated

factors influencing the response variable in opposite ways. This algorithm is related with regression.

CHAID(Kass 1980) : Later in 1980 Gordon V. Kass proposed CHAID. It is used for classification. CHAID uses chi-square test.

CART(Breiman et al. 1984) : In 1984, CART algorithm was proposed by Leo Breiman et al. CART follows the greedy search approach. It prunes the tree instead of using stopping rules. CART deals with continuous and nominal attributes data. CART can also deal with missing values. The CART algorithm is used for both classification and regression. CART uses GINI index(Classification) and sum of squared residuals(Regression).

ID3(Quinlan 1986) : The ID3 algorithm was proposed in 1986 by Ross Quinlan. The author proposed an approach to synthesize decision trees and describes ID3(Iterative Dichotomiser) in detail. ID3 is iterative in nature. ID3 restricts upto two classes for any induction task. It is used for classification. It uses entropy and information gain.

FACT(Loh, Vanichsetakul 1988) : In 1988, Loh and Vanichsetakul proposed FACT(Fast and Accurate Classification Tree) for classification. It uses LDA and ANOVA.

C4.5(Quinlan 1993) : This algorithm is an improvement on Quinlan's ID3 algorithm. It's purpose is classification. C4.5 handles both continuous and categorical data. In terms of speed C4.5 is faster than ID3. C4.5 cannot deal with missing values. It uses gain ratio. The C4.5 show biases in variable selection. It allows for more than two splits.

C5.0(Quinlan 1994) : C5.0 was proposed by Ross Quinlan in 1994. It provides several improvements on C4.5. It is used for classification and entropy is used in this algorithm. In case of speed, C5.0 is faster than C4.5. The C5.0 decision trees are significantly smaller as compared to C4.5 with matching results.

SLIQ(Supervised Learning in Quest) (Mehta, Agrawal et al. 1996) : Mehta, Agrawal and Rissanen proposed SLIQ, a decision tree algorithm that can hanble both categorical and numeric variables. By using pre-sorting technique SLIQ reduces the cost of evaluating numeric attributes. This sorting technique is combined with a breadth-first tree growing strategy to classify disk-resident datasets. Furthermore, SLIQ uses a fast subsetting algorithm for determining splits for categorical variables. SLIQ is able to scale for large data sets and classify data sets with a large number of classes, variables and examples.

QUEST(Loh, Shih 1997) : Loh and Shih in 1997 proposed QUEST(Quick, Unbiased and Efficient Statistical Tree) for classification. It uses binary splitting and is based on FACT. It avoids bias in splitting. QUEST uses F-test(non categorical variables), Chi-square test(categorical and unordered variables) and ANOVA test(ordered variables).

PUBLIC(Rastogi, Shim 2000) : In 2000, Rastogi and Shim proposed PUBLIC(PrUning and BuiLding Integrated in Classification) for classification.

CRUISE(Kim, Loh 2001) : In 2001, Kim and Loh proposed CRUISE(Classification Rule with Unbiased Interaction Selection and Estimation) for classfication. In CRUISE multiple splits are possible and it further avoids bias in splitting. It uses chi-square test.

GUIDE(Loh 2009) : In 2009, Loh proposed GUIDE(Generalized, Unbiased, Interaction, Detection and Estimation) for classification and regression. It uses binary splitting and avoids bias in splitting. It uses chi-square test.

The paper entitled "Fifty Years of Classification and Regression Trees" written in 2014 by Wei-Yin Loh has discussed the developments and research done in past 50 years in the field of trees for classification and regression. In his paper, the author has taken a comprehensive review of different algorithms.

3 Application areas

We discuss some of the application areas of recursive partitioning.

3.1 Diagnostic Analysis

This area is mostly benefitted by recursive partitioning.

Tumor classification with gene expression (Zhang et al. 2001) : As per the authors, cancer diagnosis and its treatment is dependent on exact classification of tumors. The authors analyzed a data set from the expression profiles of 2000 genes in 22 normal and 40 colon cancer tissues. The authors state that, gene expression analyses for tumor classification needs cost-effective and streamlined methodology.

Hip fracture prediction (Jin et al. 2004) : The authors propose a modified approach for an RP algorithm which they called the robust and cost saving tree(RACT). Through RACT algorithm cost-saving classification rules for hip fracture risk have been designed by the authors. The authors used 43 predictive variables to construct classfication rules. As per this algorithm, the authors grouped the predictive variables into 5 categories. The RACT algorithm uses both the learning and testing samples in the splitting step. Another modification in RACT algorithm also involves the splitting step. The authors state that, the proposed method can provide near optimum and yet economical classification rules.

Factors associated with adherence to mammography screening guidelines (Calvocoressi et al. 2005) : A study of 1229 African-American and White women in Connecticut by the authors was conducted whose commitment to mammography screening guidelines is determined over a period of 26 months in 1996-1998. The authors followed the American Cancer Society guidelines for assessing the outcome. The authors chose 22 candidate predictors to examine regarding to commitment to screening guidelines. According to the authors, overall 52.2 percent of women were found to be committed to mammography screening guidelines in the 2 years after receipt of the index screening examination. In the analysis done by the authors, CART identified six predictors of adherence to mammography screening guidelines.

Predict chemotherapy response (Rouzier et al. 2009) : As stated by the authors, the purpose of this study was to compare logistic regression model (LRM) and recursive partitioning (RP) to predict chemotherapy response in breast cancer related patients. The two different models were built having same training set of 496 patients and validated in a same validation set of 337 patients. The logistic regression model developed by the authors is based on 5 input variables, whereas the recursive partitioning model includes 4 variables.

Brain metastases (Gondhowiardjo, Aman et al. 2019) : In this study, patients with brain metastases lesions were analyzed from January 2012 until Dec. 2014. They compared three prognostic scoring tools namely recursive partitioning analysis (RPA), basic score for brain metastases (BSBM), graded prognostic assessment (GPA). Survival analysis was carried out by stratifying patients based on these three methods. RPA was the best among these three methods in stratifying patients' prognosis with brain metastases in their study population.

Identification of Diabetes with Recursive Partitioning Algorithm using Machine Learning (Lakshmi, Reddy et al. 2020) : The authors of this paper have used Pima Indian diabetes database with eight input features and one output feature. After the application of feature selection, the authors have divided the dataset into 70 percentage training data and 30 percentage test data. The authors of this paper have presented experimental evaluation with recursive partitioning algorithm for improving the accuracy in diabetes prediction at an early phase of human life. As per the authors, recursive partitioning approach is superior over prevailing approaches.

Stratification according to recursive partitioning analysis predicts outcome in newly diagnosed glioblastomas (Yang et al. 2017) : The authors undertook first of its kind study to predict the prognosis of patients with glioblastomas using recursive partitioning analysis by combining molecular parameters and clinical factors. The patients with overall survival time of less than 30 days were omitted from this study. The tree was constructed by using the variables having potential prognostic significance: age, KPS at diagnosis, six molecular markers.

Recursive Partitioning Analysis of Lymph Node Ratio in Breast Cancer Patients (Chang et al. 2015) : From the study conducted by the authors on 11349 patients, 4 prognostic factors were identified. Based on the LNR cutoffs, 4 categories are formed: very low, low, moderate, high. Then 4 risk groups starting from Class 1 with very low risk upto Class 4 with high risk were determined. RPA was used by the authors to define the breast cancer risk groups. The authors have chosen Classification and Regression Tree technique.

Risk Groups Defined by Recursive Partitioning Analysis of Patients with Colorectal Adenocarcinoma Treated with Colorectal Resection (Chang et al. 2012) : The study undertaken by the authors included 10494 patients diagnosed with colorectal adenocarcinoma who undergone colorectal resection from Taiwan Cancer Database. The timespan of this database was 2003 to 2005. For defining the risk groups, the authors have chosen classification and regression tree. The findings of this study are identification of 6 prognosis factors for cancer specific survival which results in seven terminal nodes. Also, prediction of 5 year overall and progression free survival with the help of risk groups defined by recursive partitioning analysis is done.

Classification and Regression Tree analysis to predict influenza in primary care patients (Zimmerman et al. 2016) : The authors used classification and regression trees (CART) model to evaluate probabilities of influenza. In this study, 85 per cent of the influenza positive cases were not prescribed an antiviral medication. On the other hand, 94 per cent of influenza negative cases were not prescribed an antiviral medication. The authors found that CART is good in identification of a group of individuals not requiring testing or antivirals and had predictive performance for influenza in the range from fair to good.

3.2 Insurance Sector

Insurance claims (Quan, Valdez 2018) : In this paper, multivariate tree models are applied to insurance claims data with correlated responses. In this paper, a dataset drawn from the Wisconsin Local Government Property Insurance Fund (LGPIF) is analyzed. The authors have compared the predictive performance of univariate tree-based models with multivariate tree-based models. The authors state that multivariate tree-based models are superior in performance as compared to univariate tree-based models. The paper also gives future directions to researchers to inqure into the use of multivariate decision trees in different cases of insurance claims.

3.3 Infrastructure/Civil Engineering

Deterioration prediction of bridge deck (Pittou et al. 2009) : In this paper, authors introduce binary recursive partitioning as a method of estimation of the deterioration of bridge deck. The authors have mentioned four step approach in this paper. In the first step, tree is built. In the second step tree building is stopped when : (i) single observation is found in each of the child nodes (ii) each one of the observations within each child node have the uniform distribution of predictor variables or (iii) an external limit is established by the user on the number of levels in the maximal tree. The third step is tree pruning. The final step is selecting the optimal tree. For the methodology application, the Indiana state bridge inventory database maintained by Indiana Department of Transportation (DOT) was used by the authors. The authors included bridge geometry, bridge width, the number of lanes as a set of parameters that will have impact in the BRP model. The response variable in this study was DROP which represents units of deck condition deterioration since last inspection.

3.4 Social Science/Psychology

Application of BRP to Psychology (Merkle Shaffer 2011) : This paper describes binary recursive partitioning methods in detail. The researchers have explained its use in psychological research. The researchers have mentioned that adding parameters to a model many times reduces its ability to predict future unrevealed data. The authors have also compared BRP with regression by using simulated data. By giving reference to past research by Shih(2004), the researchers have mentioned that BRP has a selection bias towards predictor variables that support a larger number of splits. The authors motivate the researchers to think about the ways in which the methods mentioned in this paper can improve psychological data analysis.

Math achivement test score (Finch 2015) : As per the author, two multilevel recursive partitioning methods Multilevel Exploratory Data Analysis (MLEDA) (Martin 2015) and Random Effects Expectation Minimization Recursive Partitioning (RE-EM tree) (Sela and Simonoff 2012) follow very different approaches to accounting for multilevel data structure when using recursive partitioning models. The author used data from the Programme for International Student Assessment (PISA) for the study. The author stated that the outcome variable of interest in this study was the math achivement test score. The author used two multilevel recursive partitioning methods. As according to the statement of the author, MLEDA offers facts concerning which predictors are maximum intently related to the outcomes, and RE-EM tree displays each specific split in the tree, with sample sizes and means for the various terminal nodes. The author suggested which method to use when on the basis of his work as well as previous research in this area.

3.5 Atmosperic Study/Meteorology

Synoptic Map-Pattern Classification using recursive partitioning and principal component analysis (Cannon et al. 2002) : The authros mentioned RPART and PCA in this paper. The authors said that for every climate detail and scenario reported in the research paper, the best recursive partitioning model outperformed the best unsupervied clustering model.

Recursive partitioning improves Paleosol proxies for rainfall (Lukens et al. 2019) : The authors reviewed and tested the role of data set size and composition on soil-climate models. The advantages of recursive partitioning stated through the authors are: subdivision of the training data set into independent groups, variable selection and data transformation. The modelling methods utilized by the authors are exponential regression, regression trees, and random forests.

3.6 Telecommunication Industry

Churn prediction(Saini, Monika et al. 2017) : The authors in this paper have focused on the loss of customers of a telecommunication company as a result of offers given by rival companies. The data set used by the researchers is an online dataset consisting of 33,000 customers (active and disconnected). The authors have used CART, QUEST, CHAID, Exhaustive CHAID in this research. The inference of this paper is Exhaustive CHAID is more efficient than others.

Customer behavior mining(Abdi et al. 2019) : The authors propose a customer behavior mining framework(CBMF). The proposed framework considers the behavior patterns of customers along with the prediction about their behavior in the future. The methods used by the authors are k-means clustering, neural network, and decision tree. The proposed framework assists the managers in implementation of CRM systems.

3.7 Pharmaceutical Sector

Drug discovery (Enade Perdana Istyastono 2015) : The author has used RPART method in this research. Use of RPART method has resulted in a significant increase in the SBVS quality to identify potent ligands for ERα.

3.8 Transporation Planning
Route choice (Deepa M.J. et al. 2012) : The authors of this paper have used RPM(Recursive Patitioning Methodology) model using DTREG software. The finding of their research is that RPM model with accuracy 71.92% is more accurate as compared to Discriminate Analysis with accuracy 67.60%.
3.9 Healthcare

Return visits of Patients' to multispecialty clinic during the 2009 H1N1 pandemic (Mgbere, Khuwaja 2009) : The authors of this paper have used nominal logistic regression and model based recursive partitioning. The decision making authorities can use finding from this study to spread awareness and understanding regarding repeated use of the healthcare system during pandemics and emergency situation handling.

Health assessments of health in Cleft Lip and/or Palate (Harrison et al. 2021) : The authors used Computerized Adaptive Testing (CAT) models and Decision Trees with binary splits. Findings of their study are: (i) Decision trees more likely made splits on the basis of patient-

reported item responses as compared to clinician-reported data, (ii) Use of binary decision trees not preferable as compared to CAT assessments for adaptive CLEFT-Q assessments.

Model-based recursive partitioning for subgroup analyses (Seibold et al. 2016) : The authors focused on detecting subgroups of patients suffering from amyotrophic lateral sclerosis(ALS). In this the subgroups differ in the treatment effect after receiving Riluzole, the approved medicine for ALS. The authors used model-based recursive partitioning to explore whether Riluzole has mostly low or high treatment effect on both functional and survival endpoints of any subgroups of patients.

Examining effects of physical activity on physical and mental health (Robinson, Gilmore 2021) : The authors have used CART method in their research. The cause of the study was to assess the effects of domain-related PA on physical and mental health. The study attempts to use CART methods to refine PA recommendations while considering domain-specific PA and discussing physical health and mental health outcomes independently.

Identifying risk profiles for childhood obesity (Hulst et al. 2015) : The authors used recursive partitioning analysis to find unique combinations of individual, familial, and neighbourhood factors that best predict obesity in children. Recursive partitioning produced 7 subgroups with a commonness of obesity equal to 8%, 11%, 26%, 28%, 41%, 60%, and 63% respectively. Findings from this study indicate that obesity is predicted by various factors in dissimilar settings and provide some indications of tending to cause obesity. The RPART method is used in this paper.

Investigating correlates of self-rated health (Nayak et al. 2017) : The authors have used classification tree analysis in this paper. This study looked into associations between several health determinants and self-rated health using recursive partitioning methods. The study used recursive partitioning in studying the combined impact of multiple health determinants.

3.10 Fashion Industry

Condition analysis and forecasting (Czekala et al. 2019) : The work described in this paper considers the problem of demand for the clothing industry's goods. It shows how this problem is linked with the mathematical problem of partitioning the set. The authors propose a solution to the problem using random partitioning. The proposed method can be a supplement to the recursive partitioning or can be used in some situations instead.

3.11 Data Imputation(dealing with missing data)

Recursive partitioning on incomplete data using surrogate decisions and multiple imputation (Hapfelmeier et al. 2011) : The authors mentioned CART, conditional inference trees, and conditional random forests in this paper. Random forests exhibited the best or at least not low performances. The overall performance of CART and conditional inference trees was also good.

3.12 Aging Study

Identify risk clusters for metabolic syndrome and its components (Pirkle et al. 2018) : The aim of the authors is to use recursive partitioning to inspect risk clusters for metabolic syndrome and its components so that to identify vulnerable populations. The study consists of three sites

form middle-income countries and two sites from high-income countries. In this paper, the study site and sex are important partitioning variables.

Correlates of loneliness in later life (Ejlskov et al. 2017) : In this paper, the authors aimd to examine the relative importance of different correlates of loneliness. The significant correlates are mental wellbeing, personal mastery, being extrovert, identifying one's partner as the closest confidant. This study shows that how recursive partitioning can be used to obtain insight regarding the relative importance of correlates of loneliness among older adults.

3.13 Drug Addiction

Examining Marijuana outcomes among college students (Wilson et al. 2018) : For this study, 8141 college students completed online survey at 11 participating universities. The sample used by the authors included 4339 lifetime users and 2129 past month users. The authors tested 3 models using RP. The first model considered 76 variables as correlates of lifetime marijuana status. The second model considered these variables as correlates of past-month marijuana status among lifetime marijuana users. The third model considered 119 variables as prospective correlates of a total score on the Marijuana Consequences Questionnaire (having marijuana use indicators) among past-month marijuana users. The authors used RP for examining correlates of marijuana-related outcomes and developed decision trees to find out important correlates of lifetime marijuana user status, past month marijuana user status and the experience of negative marijuana-related consequences.

Predicting Cannabis Abuse Screening Test(CAST) Scores (Blankers et al. 2014) : The authors have used classification tree using binary recursive partitioning in this study. An online based survey dataset covering four European countries having 2617 observations. From this, 59 potential predictors were extracted and used to partition data into subgroups with low and high risk of having a disorder related with cannabis use, on the basis of their responses on the Cannabis Abuse Screening Test. A combined model of four countries and four independent country-wise models were built. The study done by the authors show how demographic, drug use, cannabis acquisition and drug market variables can be used to form classification trees using recursive partitioning.

Risk factors associated with adolscent drug use (Tiwari et al. 2017) : In this paper, the authors have used Discriminant Analysis, Decision Tree, ANN. The research question is to examine the factors that are instrumental in illegal drug use. As per the authors the decision tree precisely classified 96.7% of the cases.

3.14 Finance Sector

Developing financial distress prediction models (Nikita Shah 2014) : The author have used methods such as Logistic Regression Model, Discriminant Analysis, Principal Component Analysis, Factor Analysis, ANN, CART, CHAID, Hybrid Model: ANN and LR, ANN and DA. The author performed test on 20 independent variables. The study's findings are limited to the corporate financial ratios of the Australian Mining Industry. As per the author, the recursive methods delivered better results as compared to the traditional univariate and multivariate analytical models.

Predicting financial distress (Gepp, Kumar 2015) : In this paper, the authors have applied Cox survival analysis model (semi-parametric) and CART decision trees (non-parametric) to financial distress predictgion. After applying these models are compared with each other and also with other popular methods. As per the authors, decision trees and survival analysis modesl have good prediction accuracy justifying their use.

3.15 Resource Utilization

Water consumption range prediction (Bermejo-Martin et al. 2021) : The authors used CART method in this paper. Based on their previous research, the author have continued their work with three different households' clusters. The authors applied CART to determine which hierarchy of variables obtained from the surveys are the important ones to predict the range of household water consumption in Huelva(Spain).

Detecting non-technical losses in power utilities (Saeed et al. 2020) : In this paper, the authors have used boosted C5.0 decision tree. The research conducted by the authors classifies loyal and fraudulent customers on the basis of the features selected. The authors have proposed an efficient energy theft detection model to find the fraudster customers in a power distribution system. The proposed model uses a set of distinct features extracted from monthly consumption data of consumers' to separate the loyal and the fraudulent customers.

3.16 Language Processing

Word sense disambiguation for Assamese (Sarmah, Sarma 2016) : Word sense disambiguation (WSD) focus on disambiguation of words having multiple sense in a context automatically. The focus of this paper is to explore research on WSD task for Assamese language using decision tree algorithm. The authors used Assamese Corpus developed in NLP laboratory of Gauhati University consisting of 1.5 million words. The authors have used C4.5 decision tree for WSD task. In this paper, the authors used Precision, Recall and F-measure as a metric for the proposed WSD method. The authors have taken initiative by using Decision Tree with Cross validation evaluation in Assamese language which is a significant contribution to Natural Language Processing field.

4 Conclusion

There are various recursive partitioning methods which can be used for prediction. Also, various areas in day-to-day life are benefitted with these methods. We have covered some of the areas of application in this paper. Recursive partitioning has a range of applications from diagnosis of disease upto churn prediction in telecommunication industry. We have summarized our observations through the tables followed.

Table 1. Recursive partitioning approaches

Algorithm	Proposed by	Type	No. of splits	Terms used
AID	Morgan, Sonquist	Regression	2	Sum of squares(explained and unexplained) by the mean
CHAID	Gordon V. Kass	Classification	>=2	Chi-square test, categorical independent variable

CART	Leo Breiman et al.	Classification, Regression	2	GINI Index(Classification), sum of squared residuals(Regression)
ID3	Ross Quinlan	Classification	>=2	Entropy, Information Gain
FACT	Loh, Vanichsetakul	Classification	2	LDA, ANOVA
C4.5	Ross Quinlan	Classification	>=2	Gain Ratio
C5.0	Ross Quinlan	Classification	>=2	Entropy
SLIQ	Mehta, Agrawal, Rissanen	Classification	2	Gini Index
QUEST	Loh, Shih	Classification	2	F-test, Chi-square test, ANOVA test
PUBLIC	Rastogi, Shim	Classification	2	Integration of the pruning with the building phase, Krichevsky and Trofimov cost function
CRUISE	Kim, Loh	Classification	Multiple	Chi-squre test
GUIDE	Loh	Classification, Regression	2	Chi-square test

Table 2. Comparative chart of ID3, C4.5, C5.0 and CART

	ID3	C4.5	C5.0	CART
Speed	Low	Faster than ID3	Highest	Average
Pruning	No pruning	Pre-pruning	Pre-pruning	Post-pruning
Missing values	Cannot deal with it	Cannot deal with it	Can deal with it	Can deal with it

References

[1] Abdi F., Abolmakarem S. : Customer Behavior Mining Framework (CBMF) using clustering and classification techniques. Journal of Industrial Engineering International, 15(Suppl 1), pp. S1-S18. (2019)

[2] Bermejo-Martin, G, Rodriguez-Monroy, C, Nunez-Guerrero, Y.M. : Water Consumption Range Prediction in Huelva's Households Using Classification and Regression Trees. Water 2021. 13, 506. https://doi.org/10.3390/w13040506 (2021)

[3] Blankers M, Frijns T, Belackova V, Rossi C, Svensson B, Trautmann F, Margriet van Laar : Predicting Cannabis Abuse Screening Test(CAST) Scores: A Recursive Partitioning Analysis Using Survey Data from Czech Republic, Italy, the Netherlands and Sweden. PLoS ONE. 9(9): e108298. Doi:10.1371/journal.pone.0108298 (2014)

[4] Breiman, L., Friedman, J.H., Olshen, R.A., Stone C.J.: Classification and Regression Trees. Chapman and Hall/CRC (1984)

[5] Calvocoressi Lisa, Stolar Marilyn, Kasl S.V.,Claus E.B.,Jones B.A. : Applying Recursive Partitioning to a Prospective Study of Factors Associated with Adherence to Mammography Screening Guidelines. American Journal of Epidemiology, 162(12), pp. 1215-1224. DOI: 10.1093/aje/kwi337 (2005)

[6] Chang, YJ., Chen, LJ., Chang, YJ., Chung KP, Lai MS : Risk groups defined by Recursive Partitioning Analysis of patients with colorectal adenocarcinoma treated with colorectal resection. BMC Med Res Methodol 12(2). https://doi.org/10.1186/1471-2288-12-2 (2012)

[7] Chang, Yao-Jen MD, PhD; Chung, Kuo-Piao PhD; Chen, Li-Ju MD; Chang, Yun-Jau MD : Recursive Partitioning Analysis of Lymph Node Ratio in Breast Cancer Patients. Medicine. 94(1). p e208 doi: 10.1097/MD.0000000000000208 (2015)

[8] Czekala M., Bukietynska A, Gurak M., Jagodzinski J, Klosowski J. : Condition Analysis and Forecasting in the Fashion Industry. International Journal of Eonomics, Finance and Management Sciences, 7(2), pp. 74-81 (2019)

[9] Deepa M.J., Alex A.P., Manju V.S.: Application of Recursive Partitioning Methodology in Route Choice. International Journal of Engineering Research & Technology, 1(10), pp. 1-9 (2012)

[10] Ejlskov Linda, Wulff Jesper, Boggild Henrik, Kuh Diana, Stafford Mai : Assessing the relative importance of correlates of loneliness in later life. Gaining insight using recursive partitioning. Aging & Mental Health, DOI:10.1080/13607863.2017.1370690 (2017)

[11] Finch W.H. : Recursive Partitioning in the Presence of Multilevel Data. General Linear Model Journal, 41(2) (2015)

[12] Gepp A., Kumar K.: Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques. Procedia Computer Science, 54(2015), pp. 396-404 (2015)

[13] Gondhowiardjo S.A.,Aman R.A., Setyawan A, Handoko, Ramli I.: Validation of recursive partitioning analysis, graded prognostic assessment and basic score for brain metastases as prognostic indices among patients with brain metastases treated with radiotherapy in Indonesia. Journal of Radiotherapy in Practice, pp. 1-5. DOI:10.1017/S1460396919000463 (2019)

[14] Hapfelmeier A, Hothorn T., Ulm K.: Recursive partitioning on incomplete data using surrogate decisions and multiple imputation. Computational Statistics and Data Analysis, 56(2012), pp. 1552-1565 (2012)

[15] Harrison C.J., Sidney-Gibbons C.J., Klassen A.F., Riff K.Y.Wong, Furniss D, Swan M.C., Rodrigues J.N.: Recursive Partitioning vs Computerized Adaptive Testing to Reduce the Burden of Health Assessments in Cleft Lip and/or Palate:Comparative Simulation Study. Journal of Medical Internet Research, 23(7) (2021)

[16] Hulst A.V., Roy-Gagnon Marie-Helene, Gauvin Lise, Kestens Yan, Henderson Melanie, Barnett T.A.: Identifying risk profiles for childhood obesity using recursive partitioning based on individual, familial, and neighborhood environment factors. International Journal of Behavioral Nutrition and Physical Activity, 12:17 DOI 10.1186/s12966-015-0175-7 (2015)

[17] Istyastono E.P.: Employing recursive partition and regression tree method to increase the quality of structure-based virtual screening in the estrogen receptor alpha ligands identification. Asian Journal of Pharmaceutical and Clinical Research. 8(6). pp. 207-210 (2015)

[18] Jin Hua, Lu Ying, Harris S.T., Black D.M.,Stone Katie, Hochberg M.C., Genant H.K.: Classification Algorithms for Hip Fracture Prediction Based on Recursive Partitioning Methods. Medical Decision Making. 24, pp. 386-398. DOI: 10.1177/0272989X04267009 (2004)

[19] Kass, G.V.: An Exploratory Technique for investigating Large Quantities of Categorical Data. Journal of the Royal Statistical Society Series C-applied Statistics, 29, pp. 119-127(1980)

[20] Kim H., Wei-Yin Loh: Classification Trees with Unbiased Multiway Splits. Journal of the American Statistical Association. 96(454). pp. 589-604 (2001)

[21] Lakshmi L., Reddy M. L., Praveen A., Suniha K.V.N. : Identification of Diabetes with Recursive Partitioning Algorithm using Machine Learning. International Journal on Emerging Technologies. 11(3), pp. 867-872 (2020)

[22] Lukens W.E., Stinchcomb G.E., Tubbs Jack : Recursive Partitioning improves paleosol proxies for rainfall. American Journal of Science. 319, pp. 819-845 (2019)

[23] Mehta M., Agrawal R., Rissanen J. : SLIQ: A fast scalable classifier for data mining. In: Apers P., Bouzeghoub M., Gardarin G. (eds) Advances in Database Technology — EDBT '96. EDBT 1996. Lecture Notes in Computer Science, vol 1057. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0014141 (1996)

[24] Merkle E.C., Shaffer V.A.: Binary recursive partitioning: Background, methods and application to psychology. British Journal of Mathematical and Statistical Psychology. 64(1). pp. 161-181 (2011)

[25] Mgbere Osaro, Khuwaja Salma : Model-Based Recursive Partitioning of Patients' Return Visits to Multispecialty Clinic During the 2009 H1N1 Pandemic Influenza(pH1N1). Online Journal of Public Health Informatics, 12(1) (2020)

[26] Morgan, J. N., Sonquist, J.A. : Problems in the Analysis of Survey Data, and a Proposal:Journal of the American Statistical Association, 58(302), pp. 415-434. http://links.jstor.org/sici?sici=01621459%28196306%2958%3A302%3C415%3APITAOS%3E2.0.CO%3B2-6 (1963)

[27] Nayak S., Hubbard A., Sidney S., Syme S.L.: A recursive partitioning approach to investigating correlates of self-rated health:The CARDIA Study. SSM-Population Health 4, pp. 178-188 (2018)

[28] Pirkle CM, Wu YY, Zuzunegui M-V, Gomez J.F.: Model-based recursive partitioning to identify risk clusters for metabolic syndrome and its components:findings from the International Mobility in Aging Study. BMJ Open. 8:e018680 (2018)

[29] Pittou M, Karlaftis M.G., Li Zongzhi : Nonparametric Binary Recursive Partitioning for Deterioration Prediction of Infrastructure Elements. Hindawi Publishing Corporation, Advances in Civil Engineering. 2009, Article ID 809767 (2009)

[30] Quan Zhiyu, Valdez E.A. : Predictive analytics of insurance claims using multivariate decision trees. Dependence Modeling. 6, pp. 377-407 (2018)

[31] Quinlan, J.R. : Induction of Decision Trees. Machine Learning, 1, pp. 81-106 (1986)

[32] Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufman. (1993)

[33] Rastogi R., Shim K.: PUBLIC: A Decision Tree Classifier that Integrates Building and Pruning. Data Mining and Knowledge Discovery. 4, pp. 315-344 (2000)

[34] Robinson, S. & Gilmore, J. : Recursive Partitioning Methods to Examine the Effects of Physical Activity on Physical and Mental Health: A Case Study at One Public University in the United States. Physical Activity and Health, 5(1), pp. 87-97. DOI:https://doi.org/10.5334/paah.98 (2021)

[35] Roman Rouzier, Charles Coutant, Benedicte Lesieur, Chafika Mazouni, Roberto Incitti, Rene Natowicz, Lajos Pusztai: Direct comparison of logistic regression and recursive partitioning to predict chemotherapy response of breast cancer based on clinical pathological variables: Breast Cancer Research and Treatment. Springer Verlag. 117(2), pp. 325-331 (2009)

[36] Salman Saeed, M., Mustafa, M. W., Sheikh, U. U., Jumani, T. A., Khan, I., Atawneh, S., & Hamadneh, N. N. : An Efficient Boosted C5.0 Decision-Tree-Based Classification Approach for Detecting Non-Technical Losses in Power Utilities. Energies. 13(12). 3242. MDPI AG. Retrieved from http://dx.doi.org/10.3390/en13123242 (2020)

[37] Sani Nisha, Monika, Garg Kanwal: Churn Prediction in Telecommunication Industry using Decision Tree. International Journal of Engineering Research & Technology. 6(4). pp. 439-443 (2017)

[38] Sarmah Jumi, Sarma S.K. : Decision Tree based Supervised Word Sense Disambiguation for Assamese. International Journal of Computer Applications. 141(1), pp. 42-48 (2016)

[39] Seibold H., Zeileis A, Hothorn T.: Model-Based Recursive Partitioning for Subgroup Analyses. Int. J. Biostat, 12(1), pp. 45-63 (2016)

[40] Shah Nikita : Developing Financial Distress Prediction Models Using Cutting Edge Recursive Partitioning Techniques: A Study of Australian Mining Performance. Review of Integrative Business & Economics Research. 3(2). pp. 103-143 (2014)

[41] Tiwari, V.K., Kumar, K., & Raj, S.: Exploration of Risk Factors Associated with Adolescent Drug Use through Cutting Edge Recursive Partitioning Techniques. Demography India:population-society-economy-environment-interactions, 46(2). pp. 81-94. http://demographyindia.in/article_document/55/779477.pdf (2017)

[42] Wei-Yin Loh : Fifty Years of Classification and Regression Trees. International Statistical Review. 82, 3, pp. 329-38. doi:10.1111/insr.12016 (2014)

[43] Wei-Yin Loh, Yu-Shan Shih : Split Selection Methods for Classification Trees. Statistica Sinica, 7(1997), pp. 815-840 (1997)

[44] Whitfield Cannon : Synoptic Map-Pattern Classification Using Recursive Partitioning and Principal Component Analysis. American Meteorological Sociey. 130. pp. 1187-1206 (2002)

[45] Wilson AD, Montes KS, Bravo AJ, Conner BT, Pearson MR Marijuana Outcomes Study Team : Making Decisions with Trees: Examining Marijuana Outcomes among College Students using Recursive Partitioning. Clin Psychol Sci. 6(5), pp. 744-754 (2018)

[46] Yang Fan, Yang Pei, Zhang Chuanbao, Wang Yongzhi, Zhang Wei, Hu Huimin, Wang Zhiliang, Qiu Xiaoguang, Jiang Tao : Stratification according to recursive partitioning analysis predicts outcome in newly diagnosed glioblastomas. Oncotarget. 8(26), pp. 42974-42982 (2017)

[47] Zhang Heping, Yu Chang-Yung, Singer Burton, Xiong Momiao: Recursive partitioning for tumor classification with gene expression microarray data. PNAS, 98(12), pp. 6730-6735 (2001)

[48] Zimmerman, R. K., Balasubramani, G. K., Nowalk, M. P., Eng, H., Urbanski, L., Jackson, M. L., Jackson, L. A., McLean, H. Q., Belongia, E. A., Monto, A. S., Malosh, R. E., Gaglani, M., Clipper, L., Flannery, B., & Wisniewski, S. R. :. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. BMC infectious diseases, 16(1), 503. https://doi.org/10.1186/s12879-016-1839-x (2016)