EFFICACY OF PART ALGORITHM IN DETERMINING RELATIONSHIP BETWEEN ECONOMIC GROWTH AND BANKING SECTOR DEVELOPMENT IN EMERGING ECONOMIES

Dr. Yasmin Shaikh
Assistant Professor (Senior Scale),
International Institute of Professional Studies, DAVV, Indore.

Dr. Manminder Singh Saluja
Assistant Professor (Senior Scale),
International Institute of Professional Studies, DAVV, Indore.

Abstract:

Machine learning methods are making their ways in explaining economic relationships. This research discusses the effectiveness of PART algorithm in determining the relationship between economic and financial variables. The study also aimed to give a comparative picture of mentioned relationship derived from PART algorithm with OneR algorithm which was conducted by authors in previous studies. Using the cross-country panel data of top ten emerging economies for the period from 1998 to 2018, the study found that performance of PART algorithm was fair as it was able to correctly classify the instances with 64.92% accuracy rate. PART algorithm succeeded in explaining the interlinkage of annual GDP growth with banking sector development, size and efficiency using categorical form of data, but its efficiency needs an improvement for highly dispersed data.

Keywords: Emerging Economies, GDP, Banking Sector, PART Algorithm

I. Introduction

Economic growth indicates expansion in the ability of an economy to produce goods and services from one period to the next. The changes in level of production are largely affected by the investments in the economy. This also attracts competition from private players. Various monetary and fiscal policies help in maintaining the balance in the economy (Darrat & Dickens, 1999). GDP is being observed as a key indicator for the economic growth by many economists therefore it is always taken as a proxy to economic growth. Economic growth further implies an increase in per capita income (Alkhazaleh, 2017).

The emerging economies incorporated in G20, forms a major part of world GDP. The G20 economies are constantly looking forward in establishing financial market stability leading to higher global economic growth. Therefore, the study of relationship between the economic growth in these emerging economies and their banking sector is much needed. An efficient banking system can transform worsening events to potential one. Emerging economies tries to venture into each available opportunity. Also escalating size and increase in population makes it difficult to for banking system to control each factor which leads to high non-performing loans and eventually it may turn into bank insolvencies, which is very costly to handle. Few countries are able to drive it into the right path and lift it up to the profitability (Caprio & Klingebiel, 1996). Since the G20 emerging economies can set up an example to the rest of the world with sound, stable & developing banking sector fueling the higher trajectory of economic growth, it becomes necessary to get new insights into this relationship for these developing economies.

II. Literature Review

There have been many researches done on the linkage between financial development and economic growth (Levine, 2000, 2005; Bonin & Wachtel, 2003; Alkhazaleh, 2017, Bongini *et.al.,* 2017, Castelli, M., 2018). Many researchers among the mentioned ones found that strong & efficient financial intermediation can boost the growth in developing economies. Credit to private sector is important as it generates employment and thus in turn contributes to economic development of a country. With proper credit and liquidity, financial imbalances can be corrected. (Malhotra, 2004). Financial deepening is important because as more liquidity is available more flexible the system can be. Most of the researches show a positive relation between broad money as a banking development indicator and GDP as proxy to economic growth (Petkovski and Kjosevski, 2014).

Financial liberalization changed the banking structure of Argentina and Brazil with well-built state owned and foreign banks. Financial institutions started diversification and technological advancement which attracted more investors and helped strengthen the domestic financial system. The entry of foreign banks was not much beneficial for Argentina to tackle with its economic crisis. While in Brazil the domestic banks welcomed the entry of foreign players and thus survived the crisis of 1994-95 (De Paula & Alves, 2007; Stanley, 2018).

Researchers argue that well-structured financial system enhance economic performance and eventually speed up economic growth. Findings shows a positive relationship for Indonesian economy (Ismail, & Pratomo, 2006). The slower growth evident in few G20 economies had been due to their financial sector problems. Though the challenges faced by Russian banks include Asset concentration, improper geographical distribution of branches and systematic weakness. The efforts were made to use financial resources wisely to meet their requirements (Pushkareva *et.al.* 2019). Studies reveal that the determinants for the bank profitability in Mexico were charges on transactions, level of capital and restraining operating outlays as well as barriers to the entry to curb competition (Rodríguez, 2015). Saudi Arabia gave financial succor to various countries. The IMF's report stated that the banking system of Saudi Arabia is well capitalized and good liquidity figures can boost growth (Malhotra, 2004).

Modern Banking in Turkey needed a robust plan to boost the economic activity through modern financial products and services for maintaining the financial stability in the economy (Nazli, 2004). Increased non-performing assets reduces the profitability of the banking sector. Less efficient and low productivity in financial system mitigate the economic growth. The global financial crisis of 2008 led to high non-performing loans in china. To promote stability in the economy, china introduced many financial reforms to upturn the economic growth (Min *et.al.* 2018). After the instigation of economic reforms in Indian banking sector, it helped in allocating resources efficiently to all the states. Although it boosted the profitability of financial institutions, but the bank credit distribution was imbalanced and less developed states faced problems. To tackle with this problem GUIDE was used to make the banking system more efficient (Arora, 2009).

Numerous researches have been done in this area that used various analytical methods in estimating this casual effect. Some researchers used the Generalized Methods of Moments (GMM) dynamic panel method to investigate cross countries' relationship between banking sector development and economic growth (Petkovski and Kjosevski, 2014; Bongini *et.al.*, 2017) while other researches like that of Alkhazaleh (2017), used ordinary least square regression methods to find a positive relation between bank performance indicators and GDP. To overcome the problems associated with these methods of analyzing required relationship, machine learning tools are used which gives precise outcomes, and study by generating logical algorithms & models and thus help in better decision making (Sutskever *et.al.* 2013).

A recent study conducted by Saluja *et.al.* (2020) investigated the relationship between banking sector development and economic growth using OneR algorithm. The research generated one rule respectively for each banking attribute taken individually & jointly, as independent variable with seven other economic attributes. The precision rate for their study was above 60%. The classification algorithms namely Naïve bayes, BayesNet, PART, JRip and OneR were applied on Hypothyroid health database for the purpose of finding better techniques for classification (Parsania V. *et. al.*, 2014). The result revealed that, if precision, specificity, false positive rate, F-Measure and correctly classified instances are the key parameters to be considered for the classification; PART is preferable to apply.

The above discussed literature clearly states that the gap exists in analysis of economic variables using machine learning methods specially with PART algorithm. Machine learning methods are expected to give better predictability compared to various other statistical methods. The present study therefore will aim to determine the relationship between various economic and financial series using PART algorithm. The objective is to establish the efficacy of PART algorithm in estimation of relationship between economic and financial variables.

III. Methodology & Data

In this study, PART classifier is used to generate decision list. PART is a separate-and-conquer rule learner. PART creates a decision tree and then generates a rule. A tree is built for the current set of instances. The rule for largest leaf is read and the sub tree is pruned. The instances that were covered by the rule being generated are then removed. The procedure is repeated for remaining instances. A partial tree is built instead of a full one, therefore the algorithm is named as PART algorithm. The algorithm produces set of rules called as "decision lists" which are planned set of rules. The rules in the decision list are interpreted in sequence. The data records that are classified in first rule are eliminated from dataset and next rule is generated for remaining data records. The process continues till all data records are classified.

This rule generation was done using Weka (Frank *et.al.*, 2016). There were four banking attributes and seven economic attributes that were taken together to determine the decision list. The target attribute in the data under analysis is annual GDP growth rate [GDP Rate].

Data Selection

The current research aims to test the effectiveness of PART algorithm and also provide a comparative picture of its performance to that of OneR algorithm. Thus, the present research continues with its study on same set of countries and variables as were taken by Saluja *et.al*. (2020) in their most recent research. However, the present research used the panel data of ten major emerging economies of G20 for twenty years from 1998 to 2018. The variables used by Saluja *et.al.* (2020) are reproduced hereby for better understanding of the results. The present study also used four banking sector variables namely Annual Broad Money Growth Rate *[Broad Money]*, Domestic credit to private sector by banks as % of GDP *[Domestic Credit]*, Interest rate spread (lending rate minus deposit rate) *[Interest Rate Spread]*, Bank's Non-Performing loans to total gross loans (%) *[Non-Performing Loan]*. The economic attributes are Natural Log of GDP per capita *[LGDPPC]*, Exports of goods and services (% of GDP) *[Exp Goods Service]*, Consumer price index is taken as indicator of Inflation Rate *[Inflation Rate]*, Govt. Spending measured as General government final consumption expenditure as % of GDP) *[Govt Spending]*, Investments taken as Gross capital formation as % of GDP *[Investments]*, Rule of Law: Estimate *[Rule Estimate]*, Annual GDP growth rate *[GDP Rate]*. The names in the parenthesis are the representative names of the variables taken for analysis. A representative sample of panel data of ten emerging economies that was collected from World Development Indicator (WDI) – a database of World Bank, is shown in Table below.

Table - 1 Random Sample Data of China and India among Top Ten Emerging Economies on Economic and Monetary Variables

Year	Broad Money Growth	Interest Rate Spread	Domestic Credit	Non-Performing Loans	LGDP Per Capita	Exports	Inflation Rate	Govt. Spending	Investments	Rule of Law	GDP Annual Growth Rate
China 1998	14.904	2.610	105.071	48.600	6.720	18.342	81.836	14.835	35.679	-0.457	7.838
1999	14.666	3.600	110.271	32.900	6.772	18.163	80.689	16.239	34.965	-0.457	7.667
2000	12.325	3.600	111.013	22.400	6.866	20.893	80.970	16.633	34.430	-0.527	8.492
2001	15.042	3.600	109.868	29.800	6.960	20.312	81.552	16.090	36.422	-0.527	8.340
2002	13.140	3.330	117.329	26.000	7.046	22.645	80.955	15.603	37.078	-0.498	9.131
2003	19.240	3.330	125.521	20.400	7.161	26.981	81.868	14.677	40.632	-0.530	10.036
2004	14.887	3.330	118.552	13.200	7.319	31.061	84.999	13.902	42.894	-0.529	10.111
2005	16.742	3.330	111.807	8.600	7.469	33.830	86.509	13.995	41.391	-0.589	11.396
2006	22.116	3.600	109.128	7.100	7.649	36.035	87.936	13.949	40.933	-0.639	12.719
2007	16.736	3.330	105.764	6.200	7.899	35.435	92.172	13.490	41.485	-0.542	14.231
India 2008	20.495	3.141	49.559	2.400	6.906	24.097	80.531	10.538	37.851	0.086	3.087
2009	17.996	2.731	48.124	2.200	7.005	20.401	89.294	11.460	40.112	0.014	7.862
2010	17.802	3.085	50.555	2.386	7.213	22.401	100.000	11.008	40.220	-0.037	8.498
2011	16.138	3.349	51.289	2.670	7.285	24.540	108.858	11.084	39.590	-0.091	5.241

2012	11.046	3.068	51.889	3.374	7.275	24.534	118.996	10.684	38.347	-0.072	5.456
2013	14.832	3.011	52.386	4.028	7.279	25.431	131.975	10.295	34.023	-0.057	6.386
2014	10.587	2.850	51.882	4.346	7.361	22.968	140.360	10.441	34.268	-0.063	7.410
2015	10.618	2.900	51.868	5.884	7.381	19.813	148.603	10.428	32.117	-0.047	7.996
2016	6.801	2.842	49.195	9.186	7.455	19.195	155.945	10.306	30.212	-0.029	8.170
2017	10.431	3.013	48.780	9.980	7.591	18.781	159.829	11.030	30.941	0.005	7.168
2018	10.520	3.550	50.046	9.461	7.606	19.738	167.598	11.229	31.308	0.026	6.811

Source: - **WDI, World Bank**

Data Preprocessing

The decision list from the PART algorithm will be estimated using the data stated in table 1. However, since the decision estimation expected from the said algorithm is categorical, all the 210 observations for all Eleven variables have to be categorized. The categorization of the variables was done in three fields namely *Increasing, Decreasing and No Change*; similar to the categorization done by Saluja *et.al.* (2020). Table 2 below presents the sample of the categorical data that was converted from numeric sample data shown in table 1.

Table - 2 Data Categorization of Sample Data

Broad Money Growth	Interest Rate Spread	Domestic Credit	Non-Performing Loans	LGDP Per Capita	Exports	Inflation Rate	Govt. Spending	Investments	Rule of Law	GDP Annual Growth Rate
Decreasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Increasing	Decreasing	Increasing	Decreasing
Decreasing	Increasing	Increasing	Decreasing	Increasing	Decreasing	Decreasing	Increasing	Decreasing	Decreasing	Decreasing
Decreasing	Decreasing	Increasing	Decreasing	Increasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Increasing
Increasing	Decreasing	Decreasing	Increasing	Increasing	Decreasing	Increasing	Decreasing	Increasing	Decreasing	Decreasing
Decreasing	Decreasing	Increasing	Decreasing	Increasing	Increasing	Decreasing	Decreasing	Increasing	Increasing	Increasing
Increasing	Decreasing	Increasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Increasing	Decreasing	Increasing
Decreasing	Decreasing	Decreasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Increasing	Increasing	Increasing
Increasing	Decreasing	Decreasing	Decreasing	Increasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Increasing
Increasing	Increasing	Decreasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Decreasing	Increasing
Decreasing	Decreasing	Decreasing	Decreasing	Increasing	Decreasing	Increasing	Decreasing	Increasing	Increasing	Increasing
Decreasing	Decreasing	Increasing	Decreasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Decreasing
Decreasing	Decreasing	Decreasing	Decreasing	Increasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Increasing
Decreasing	Increasing	Increasing	Increasing	Increasing	Increasing	Increasing	Decreasing	Increasing	Decreasing	Increasing
Decreasing	Increasing	Increasing	Increasing	Increasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Decreasing
Decreasing	Decreasing	Increasing	Increasing	Decreasing	Decreasing	Increasing	Decreasing	Decreasing	Increasing	Increasing
Increasing	Decreasing	Increasing	Increasing	Increasing	Increasing	Increasing	Decreasing	Decreasing	Increasing	Increasing
Decreasing	Decreasing	Decreasing	Increasing	Increasing	Decreasing	Increasing	Increasing	Increasing	Decreasing	Increasing
Increasing	Increasing	Decreasing	Increasing	Increasing	Decreasing	Increasing	Decreasing	Decreasing	Increasing	Increasing
Decreasing	Decreasing	Decreasing	Increasing	Increasing	Decreasing	Increasing	Decreasing	Decreasing	Increasing	Increasing
Increasing	Increasing	Decreasing	Increasing	Increasing	Decreasing	Increasing	Increasing	Increasing	Increasing	Decreasing
Increasing	Increasing	Increasing	Decreasing	Increasing	Increasing	Increasing	Increasing	Increasing	Increasing	Decreasing

PART algorithm was applied to the preprocessed data with classifier options as shown in table below.

Table – 3 Classifier Options

Classifier Option	Value
Confidence threshold for pruning	0.25
Minimum number of objects per leaf	6
Reduced error pruning	False
Use binary splits only	False
Generate unpruned decision list	False
Seed for random data shuffling	1

IV. Results & Discussion

PART is an algorithm for inferring rules by repeatedly generating partial decision trees. It combines the two major paradigms for rule generation – creating rules from decision trees and the separate-and-conquer rule-learning technique (Frank & Witten, 1998). The experiments on standard datasets show that it produces rule sets that are as accurate as and of similar size to those generated by C4.5, and more accurate than RIPPER's. Moreover, it operates efficiently, and because it avoids post-processing, it does not suffer from the extremely slow performance on pathological example sets for which the C4.5 method has been criticized. A new data is compared to each rule in the list in turn, and the item is assigned the class of the first matching rule. PART builds a partial C4.5 decision tree in each iteration and makes the “best” leaf into a rule.

Performance Evaluation: PART Algorithm

Table 4 shows the values of various performance measures for results of PART classifier. When PART classifier is executed with all the banking attributes along with all economic attributes, the percentage of accuracy is 64.9289 %. The Kappa statistic value is 0.2926 which implies that classifier is fairly classifying the data. The weighted average values of Precision and F-measure for all datasets is more than 0.6. It implies that the accuracy level is high and the rules can predict the value of target attribute fairly accurately.

Table - 4 PART Algorithm Evaluation Parameters

Measure	Value	Performance Level
Correctly Classified Instances	64.9289 %	Fair
Kappa Statistics	0.2926	Fair
Mean Absolute Error	0.423	Fair
Root Mean Squared Error	0.4857	Fair
Precision (Weighted Average)	0.648	Fair
Recall (Weighted Average)	0.649	Fair
F-measure (Weighted Average)	0.648	Fair

Rule Sets with PART algorithm:

OneR algorithm respectively found Govt Spending and Bank NPL, as dominant predictor variables while analyzing the impact of banking sector development on economic growth taking each banking sector variable individually and jointly along with other economic variables (Saluja *et.al.*, 2020). To further gain insight into the impact of banking sector development, size and efficiency on economic attributes, PART algorithm is applied. On applying PART algorithm on different datasets under analysis with classifier options as mentioned in table 3 using 10-folds cross validation the study derived a decision list with seven rules as shown in table 5.

Table – 5 Relationship Between Economic Growth and Banking Sector Development

PART Algorithm Decision List:	
Rule 1: Bank NPL is Increasing & Interest Rate Spread is Increasing	→ GDP is Decreasing (40.0/8.0)
Rule 2: Govt. Spending is Decreasing	→ GDP is Increasing (78.0/25.0)
Rule 3: Rule of Law is Increasing, Bank NPL is Decreasing & Domestic Credit is Increasing	→ GDP is Decreasing (20.0/5.0)
Rule 4: Bank NPL is Decreasing & Rule of Law is Decreasing	→ GDP is Increasing (29.0/10.0)
Rule 5: Bank NPL is Increasing	→ GDP is Decreasing (24.0/6.0)

Rule 6: Rule of Law is No Change	→ GDP is Decreasing (10.0/1.0)
Rule 7:	→ GDP is Increasing (10.0/3.0)

The rules in the decision list are interpreted in sequence. The data records that are classified in first rule are eliminated from dataset and next rule is generated for remaining data records. The process continues till all data records are classified. In the above Decision List, the first rule indicates that when bank's non-performing loans and interest rate spread both are increasing for all the ten emerging economies together then it is putting a negative impact on their annual GDP growth rate. Here 40 data records are classified out of which 32 are correctly classified. Rule 2 classified 78 out of 210 records with 53 correctly classified instances. This rule stated a negative relationship between annual GDP growth rate and govt. spending in these emerging economies. The results of the PART algorithm show that Rule 1 and Rule 2 together makes up biggest share of records and that too correctly classified ones. Both rules go in line with text book relationships that higher Govt. spending by crowding out private credit puts a negative impact on investment, production and employment, thus leading to a lower GDP growth rate and vice versa. Higher level of NPL and increase in interest rate spread reduces the efficiency of banking sector through its lending capacity thereby passing the contagion on to GDP growth rate. Rule 3 to 6 also reiterates the above generalization of hypothesis.

V. Conclusion

Data mining is slowly taking a prominent role in determining relationship among economic and financial variables. The better predictability of machine learning methods is making them the flavour of the day. The present study also made an attempt to fill a major gap in literature by testing the efficacy of PART algorithm - one among classification methods of machine learning- on economic and financial relationships. The study established that PART algorithm was fairly successful in determining the relationship between economic growth and banking sector development using panel data of top 10 emerging economies among G20 countries. However, it can be concluded that the higher dispersion that persist in data of these nations may had been a major reason for fair accuracy rather than a high accuracy level of this algorithm. Researchers can work on creating an improved PART algorithm in future researches which can perform more accurately with wider heterogeneous data also.

References

1. Adriansson, N.; and Mattsson, I.; (2015), Forecasting GDP Growth, or How Can Random Forests Improve Predictions in Economics?, Uppsala University Department of Statistics, February.
2. Alkhazaleh, A. M. K. (2017), Does Banking Sector Performance Promote Economic Growth? Case Study of Jordanian Commercial Banks, Problems and Perspectives in Management, LLC, CPC Business Perspectives.
3. Arora, R. U. (2009), Bank Credit and Economic Development: An Empirical Analysis of Indian States, Journal of Asian Public Policy, 2 (1), 85–104.
4. Bongini, P.; Iwanicz-Drozdowska, M.; Smaga, P.; & Witkowski, B. (2017), Financial Development and Economic Growth: The Role of Foreign-Owned Banks in CESEE Countries, Sustainability, 9 (3), 335.
5. Bonin, J., and Wachtel, P. (2003), Financial Sector Development in Transition Economies: Lessons from the First Decade. Financial Markets, Institutions and Instruments. Blackwell Publishing Inc.
6. Breiman, L., (2001), Random Forests, Machine Learning, 45 (1), 5-32.
7. Caprio, G.; and Klingebiel, D.; (1996), Bank Insolvencies: Cross Country Experience. The World Bank Policy Research.
8. Castelli, Maddalena (2018), Financial Sector Development and Economic Growth: Revisiting the Relationship, Perspectives: A Publication of International Development Program, School of Advanced International Studies, John Hopkins University, April 12.
9. Darrat, A. F.; and Dickens, R. N.; (1999), On the Interrelationships Among Real, Monetary, and Financial Variables. Applied Financial Economics, 9 (3), 289–293.
10. De Paula, L. F.; and Alves, A. J.; (2007), The Determinants and Effects of Foreign Bank Entry in Argentina and Brazil: A Comparative Analysis. Investigacion Economica, 66 (259), 63–102.
11. Frank, E.; Hall, M. A.; Witten, I. H. (2016), The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Fourth Edition.
12. Frank E. and Witten I. H. (1998), Generating Accurate Rule Sets Without Global Optimization, Fifteenth International Conference on Machine Learning, 144-151.

13. Ismail, A. G.; and Pratomo, W.; (2006), Financial Intermediation and The Real Sector in Indonesia, Savings and Development, 30 (2), 131–142.
14. Levine, R.; Loayza, N.; and Beck, T.; (2000), Financial Intermediation and Growth: Causality and Causes. Journal of Monetary Economics, 46 (1), 31–77.
15. Levine, R. (2005), Finance and Growth: Theory and Evidence, Handbook of Economic Growth, Chapter 12, 1, Part A, 865-934, Elsevier.
16. Malhotra, R. (2004), Saudi Arabia, Practical Neurology, Blackwell Publishing Ltd.
17. Min, Z.; Weidong, C.; Jingtong, Z.; Xinzhe, G.; & Qiyue, X.; (2018), The Development of China's Financial System: A Global Perspective, China Economic Journal, 11 (1), 25–43.
18. Nazli, H. K.; (2004), Banking on Turkey, National Interest, (78), 114.
19. Parsania, Vaishali; Jani, Nalin; Bhalodiya, Navneet; (2014), Applying Naïve Bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis, IJDI-ERET, 3 (1), 60-64.
20. Petkovski, M., and Kjosevski, J. (2014). Does Banking Sector Development Promote Economic Growth? An Empirical Analysis for Selected Countries in Central and South Eastern Europe. Economic Research-Ekonomska Istrazivanja, 27 (1), 55–66.
21. Pushkareva, L. V.; Galochkina, O. A.; and Bezgacheva, O. L.; (2019), Current Trends in the Banking System of Russia, Espacios, 40 (4).
22. Rodríguez, R. Chavarín; (2015), Determinants of Commercial Bank Profitability in Mexico. Econoquantum, 12 (1), 97–123.
23. Saluja, Manminder Singh; Shaikh, Yasmin; and Kamle, Pranita; (2020), Does Banking Sector Development Promote Economic Growth? An Empirical Analysis for Emerging Economies Using One Rule Algorithm, Shodh Sarita, 7 (26).
24. Stanley, E.; (2018), Emerging Market Economies and Financial Globalization, Part-1, The Financial Globalization Journey: The General Framework.
25. Sutskever, I.; Martens, J.; Dahl, G.; and Hinton, G.; (2013), On the Importance of Initialization and Momentum in Deep Learning, International Machine Learning Society 30th International Conference on Machine Learning, 2176–2184.