Empirical and Indian Automotive Equity Portfolio Decision Support

P. Sankar, P. James Daniel Paul, Siddhant Sahu

***Abstract*—**A brief review of the empirical studies on the methodology of the stock market decision support would indicate that they are at a threshold of validating the accuracy of the traditional and the fuzzy, artificial neural network and the decision trees. Many researchers have been attempting to compare these models using various data sets worldwide. However, the research community is on the way to the conclusive confidence in the emerged models. This paper attempts to use the automotive sector stock prices from National Stock Exchange (NSE), India and analyze them for the intra-sectorial support for stock market decisions. The study identifies the significant variables and their lags which affect the price of the stocks using OLS analysis and decision tree classifiers.

***Keywords*—**Indian Automotive Sector, Stock Market Decisions, Equity Portfolio Analysis, Decision Tree Classifiers, Statistical Data Analysis.

I. Introduction

STOCK market decisions are dynamic in intra-day. But if there is an opportunity to see the buy or sell decisions a few days ahead is always the desired objective of the analyst in the markets. This paper attempts to identify the determinants through the traditional models and the computational decision models. The regression models have been modified into non-linear models and two types of the decision trees formed using machine learning algorithms have been used to analyze the data. This study covers the data for the last calendar year obtained for 8 automotive sector companies from the NSE historical data.

II. Review of Literature

Jar-Long Wang et al. (2006) [1], in their research paper stated that the accuracy of the forecasts are determined by comparing each individual’s test case prediction with its actual outcome on a percentage basis, and the return rates is determined by buy-and-hold for 100 trading days. They use variables like stock price data, upward class, downward class, buy class and not to buy class. They use two-layer bias decision tree. They conclude with a comparison of random purchases, the results indicate the system presented here not only has excellent out-of-sample forecasting performance, but also delivers a significant improvement in investment returns for all listed companies.

Chih-Fong Tsai et al. (2010) [2], in their article, used variables like US gross national income, US producer price index, US annual changes in consumer price index, US personal consumption expenditures, US annual changes in industrial production index, US current account to GDP ratio, Taiwan unemployment rate, quasi money, export amount to US, US merchandise trade volume, export order for electric products, GNP deflator, US monetary supply, narrow monetary supply and subjected these variables to principle component analysis, genetic algorithm and decision tree. They concluded that the intersection between PCA and GA and the multi-intersection of PCA, GA, and CART perform the best, providing the highest rate of prediction accuracy and the lowest error rate of predicting stocks' rise.

Tomer Geva et al. (2014) [3] whilst, conducting an empirical evaluation of an automated intraday stock recommendation system incorporated both market data and textual news utilize overall, 51,263 news items. They calibrated sentiment scores using models like Neural network (NN), Decision tree involving a genetic algorithm and stepwise logistic regression. This study showed that integrating market data with textual data contributes to improving the modeling performance and that using more advanced textual data representations further improves predictive accuracy. However, these results strongly depend on the joint selection of both data representation and forecasting algorithm

Man Hong Wong et al. (2014) [4] in an article used models like probability model, derived conditional value-at-risk, single cluster model, numerical algorithm, probability model and the downside risk model They concluded that no neater and simpler form is achieved, which implies we will have to rely mainly on numerical methods.

Muh-Cherng Wu et al. (2006) [5] in their research work used two stock markets data, Taiwan and NASDAQ, analyzed variables like number of trading points and percentage of trading points with positive return using decision tree algorithm (C4.5). They conclude that empirical tests reveals that the filter rule performs the best at (n, k)Z (10, 10%) in both the markets. The proposed trading method outperforms Lin’s method, substantially in NASDAQ market and slightly in Taiwan.

Robert K. Lai at al. (2009) [6] in their article used stock trading data from 2005 to 2005 on TSEC (Taiwan Stock Exchange Corporation) on variables like capital stock, revenue situation, EPS, turnover number, net worth and market value ratio, price-earnings ratio, six days moving average, six days

P. Sankar is a doctoral scholar at the VIT Business School, VIT University, Chennai, Tamil Nadu-600127, India (e-mail: sankar.p2013@vit.ac.in).

P James Daniel Paul is a professor of economics and finance at VIT Business School, VIT University, Chennai, Tamil Nadu-600127, India (phone: +91 9840294590; e-mail: jamesdanielpaul.p@vit.ac.in).

Siddhant Sahu is a senior under-graduate student at the School of Mechanical and Building Sciences, VIT University, Chennai, Tamil Nadu 600127,India (phone: +91 8939590061; e-mail: siddhant.sahu2011@vit.ac.in).

bias, six days relative strength index, nine days stochastic line, moving average convergence and divergence, 13 days psychological line, volume, buy, sell data in data clustering technique, a fuzzy decision tree (FDT) and genetic algorithms (GA). They concluded that investors prefer buy or sell advice rather than the price forecast. This can be achieved by effective data clustering methods, a different data mining model and different data fossilization methods.

Il Suh Son et al. (2009) [7] attempt to develop an early warning system for global institutional investors at emerging stock markets based on machine learning forecasting. Classifiers were built on an 'if, then, else' algorithm. In this study, the EWSGII is proposed which forecasts the movements of GII by classifying the future market condition. For this, the oracle and trained classifiers were introduced.

Pei-Chann et al. (2011) [8] ventured in a trend discovery in financial time series data using a case based fuzzy decision tree. This forecasting model integrates a data clustering technique, a fuzzy decision tree (FDT), and genetic algorithms (GA) to construct a decision making system based on historical data and technical indexes. After using different input factors and different forecasting models, such as CART and C4.5, possible candidate models for improving the accuracy of stock movement prediction, they concluded that researchers can use different kinds of fuzzy membership functions to transform the original data, including trapezoid membership functions and gauss membership functions.

Wen-Shiung Lee et al. (2011) [9] analyzed decision making factors for equity investment by DEMATEL and analytic network process using fifteen questionnaires in a survey conducted between October and December, 2008 on a 7 sample stocks, 10 month data. They used fundamental analysis, technical analysis, and institutional investor analysis and finally adopt the methods of DEMATEL and ANP to analyze the interdependences between key factors of stock investment decision making.

David Diaz et al. (2011) [10] in their analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices, used the COMPUSTAT database (Standard and Poor's Compustat Resource Center, 2009) to provide supplementary profiling financial information about the selected cases, such as the SIC Code, market capitalization and beta. They use variables like ZO1 that refers to the returns indicator, ZAR1 to the abnormal returns indicator, ZS1 to the liquidity indicator and RV3 to the volatility indicator. They use the regression and frequency of outlier's analysis, confusion matrix, decision trees and conclude that when returns are within normal ranges, isolated jumps in liquidity are associated with suspicious trades in more than 20% of the cases

Tsung-Sheng Chang (2011) [11] in a comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction, used 10 different stocks in 320 data sets to study variables like current day closing price of a stock, previous day closing price, OTC index, stock ID in artificial neural networks (ANN), decision trees and the hybrid model of ANN and decision trees (hybrid model). They concluded that the average accuracy of ANN is 15.31%, the highest, in terms of match with real market stock prices, followed by decision trees, at 14.06%; hybrid model is 13.75%.

Wangren Qiu et al. (2012) [12] while forecasting Shanghai composite index based on fuzzy time series and improved C-fuzzy decision trees used Shanghai composite index over a ten-year period using C-fuzzy decision tree WCDT. They proposed a new method for fuzzy time series forecasting based on weighted C-fuzzy decision trees which can obtain more stable results with lower computational cost.

Shu-Hsien Liao et al. (2013) [13] investigated data mining and co-movements on the Taiwan and China stock markets for future investment portfolio using indices of 30 categories from Hong Kong Stock Exchange (HKEX) and Shanghai stock exchange (SSE) with a total of 795 transaction days. This study considered that a stock market has strong associations with both inside and outside factors.

Chih-Fong Tsaia et al. (2011) [14] while predicting stock returns by classifier ensembles on the Taiwan Economic Journal (TEJ) dataset, from the second quarter of 2002 to the third quarter of 2006 used variables like capital structure, debt ratio, long-term capital, amortization capability, current ratio, quick ratio, interest cover, business operation capability, total asset turnover ratio, fixed asset turnover ratio, inventory turnover ratio, accounts receivable turnover ratio, profitability return on assets, margin before interest and tax, net assets per stock, return on stockholder's equity, cash flows, cash flow ratio, others constant net assets growth ratio, net assets growth ratio after tax, frequent interest growth ratio after tax, return on total assets growth ratio, return ratio of the last quarter, economic indicators, deposit interest rate, currency transferring rate (US dollars to Taiwan dollars), discount rate, money supply, consumer price index, wholesale price index, unemployment rate, bond trading amount, total assets of listed companies, Taiwan stock index and industrial production index in single classifiers, multi-layer perception (MLP) neural network, classification and regressing tree (CART) decision trees, and logistic regression (LR). They state that the homogeneous classifier ensembles by majority voting are particularly good at predicting positive returns, while the performance of predicting negative returns is better than the single best MLP model.

Preeti Paranjape et al. (2013) [15] in a stock market portfolio recommender system based on association rule mining for BSE-30 sensitive Index, the S&P CNX Nifty or NSE-50, S&P CNX-100 and DOW-30 Industrial Average with a lag of 2 days, use variables like stock name, price, value in an association rule mining (ARM). They infer that the application of soft computing techniques like ARM and fuzzy classification in the design of an efficient recommender system.

Agnes Virlics (2013) [16] in a study on investment decision making and risk, surveyed extensive literature on investment decisions in the economic theory, investments and risk and decision making and risk a behavioral and neuro-economic approach and concluded that investments, in most cases, risk and uncertainty is subjectively perceived and it involves psychological and emotional factors.

Based on the survey of literature, the Indian automotive sector equity prices data set is analyzed.

III. Data and Analysis Methodology

The data considered for analysis in this study is the daily stock price data of 8 major Indian automotive sector firms namely Ashok Leyland, Bajaj Auto, Eicher Motors, Hero Motors, Hindustan Motors, Mahindra & Mahindra, Maruti Suzuki India, Tata Motors. The data has been collected from 8th January, 2013 to 9th January, 2014 i.e. for a year and contains data for 250 trading days. This historical data was collected from the official website of NSE. The major variables considered for characterization of the stock are date, closing price (in rupees) of the stock for the day and total traded quantity of stocks on a specific day. Apart from these variables, 10 consecutive time lag variables have also been introduced for 'close price' and 'total traded quantity' for each automotive stock for the analysis. The company names like Ashok Leyland and Hindustan Motors Limited have been commonly abbreviated as 'AL' and 'HML' respectively in the variables.

In order to understand the data characteristics better, the central tendencies, deviations and variance of the data have been analyzed in Table I.

TABLE I
DATA DESCRIPTIVE STATISTICS

	Mean	Std. Deviation	Variance
AL Close Price	19.1754	4.200	17.64347876
AL Quantity	6783753.29	5294260.66	2.80292E+13
Bajaj Close Price	1925.9846	120.271799	14465.30553
Bajaj Quantity	394467.804	247966.314	61487293068
Eicher Close Price	3497.613	689.088984	474843.6279
Eicher Quantity	23455.592	25752.052	663168182.5
Hero Close Price	1829.0576	196.287385	38528.73763
Hero Quantity	343865.892	245495.181	60267883874
HML Close Price	8.3754	1.18305678	1.399623333
HML Quantity	249956.96	414539.53	1.71843E+11
Mahindra Close Price	901.3642	52.4146153	2747.291896
Mahindra Quantity	1236427.19	643992.443	4.14726E+11
Maruti Close Price	1520.1288	145.618614	21204.78079
Maruti Quantity	693449.14	477811.38	2.28304E+11
Tata Close Price	167.8608	20.379936	415.3417905
Tata Quantity	2131173.31	1220997.4	1.49083E+12

Two main methods have been used for the analysis namely regression analysis and decision tree classifiers. Initially, a stepwise regression analysis has been done to obtain various models with different variables. This stepwise analysis was done separately for each stock price by considering it as a dependent variable and the rest as independent variables. For each analysis, the model with the maximum number of variables with the best fit has been considered. The variables of that model along with the exponential log of total traded quantity have been subject to an enter regression analysis to obtain the individual correlation coefficients and their respective t values.

In the second method, the data was first converted into a multi-class problem wherein separate analysis was done on every stock by converting that specific stock price variable into a class variable. The mean closing price was taken as the classifying parameter. Any price value above the mean price was assigned the class 'Sell' and any value below the mean was assigned the class 'Buy'. Two decision tree classifiers were used namely 'J48 Decision Tree' and 'Random Decision Tree' and the respective trees signifying the classification rules and significant variables were obtained. A ten-fold cross validation was also performed in order to compute the classification accuracy of the classifiers in order to evaluate the effectiveness of the classifiers in classifying data.

IV. Results and Discussion

A. Regression Analysis

The current closing price of the respective stock is the dependent variable in the simple OLS model. The results for regression analysis for each stock price have been presented in Tables II to X.

TABLE II
REGRESSION ANALYSIS OF ***ASHOK LEYLAND***

Model	Unstandardized Coefficients		t
	B	Std. Error	
(Constant)	-0.56798	0.814676	-0.69719
AL Price LAG1	0.854033	0.024447	34.93429
Tata Quantity LAG2	5.42E-08	2.02E-08	2.678936
Mahindra Close Price	0.006183	0.001407	4.393491
Mahindra Price LAG1	-0.00675	0.001844	-3.66249
Eicher Quantity LAG1	-3E-06	8.58E-07	-3.46083
Eicher Quantity LAG4	1.72E-06	8.75E-07	1.964395
Hero Quantity LAG8	3.44E-07	9.83E-08	3.498967
HML Close Price	0.538472	0.094445	5.701458
Hero Price LAG10	-0.00154	0.000242	-6.38897
Mahindra Price LAG7	0.004729	0.000915	5.171326
Eicher Quantity LAG8	-2.2E-06	8.42E-07	-2.56342
Bajaj Quantity LAG5	-2.6E-07	9.08E-08	-2.8769
Bajaj Quantity LAG2	3.52E-07	9.47E-08	3.723032
Mahindra Price LAG4	-0.00362	0.00126	-2.87685
Hero Quantity LAG7	-2.4E-07	9.31E-08	-2.60259
Maruti Quantity LAG6	1.15E-07	4.67E-08	2.46543
Tata Quantity LAG3	4.89E-08	2.03E-08	2.415058
Bajaj Quantity LAG1	-2.6E-07	9.01E-08	-2.84334
HML Price LAG1	-0.23728	0.100761	-2.35491
MahindracPriceLAG2	0.003265	0.001641	1.990097
AL Quantity LN	0.006306	0.039389	0.160102

In the case of Ashok Leyland share prices in NSE, 20 explanatory automotive sector variables indicate significant causality.

TABLE III
REGRESSION ANALYSIS OF ***BAJAJ AUTO***

Model	Unstandardized Coefficients		T
	B	Std. Error	
(Constant)	153.67767	58.622673	2.6214716
Bajaj Price LAG1	0.9163661	0.0243677	37.605697
Tata Close Price	2.2137689	0.5160243	4.2900481
Tata Price LAG1	-1.700565	0.4947892	-3.4369484
Maruti Price LAG9	-0.0154367	0.0233871	-0.6600514
AL Quantity LAG3	1.118E-06	4.432E-07	2.5216716
HML Close Price	21.572082	5.2601942	4.101005
HMM Price LAG2	-12.434684	5.9089564	-2.104379
Eicher Price LAG9	-0.012242	0.0067448	-1.8150209
HML Price LAG8	-12.116723	4.1842782	-2.8957738
Maruti Quantity LAG9	-9.106E-06	3.725E-06	-2.4448059
Maruti Quantity LAG10	5.657E-06	3.639E-06	1.5545426
Hero Close Price	0.2713401	0.0575927	4.7113648
Hero Price LAG1	-0.246755	0.0573613	-4.3017712
Bajaj Quantity LN	-2.9065289	3.4936547	-0.8319451

In the case of Bajaj Auto, share prices in NSE, 13 explanatory automotive sector variables indicate significant causality.

TABLE IV
REGRESSION ANALYSIS OF ***EICHER***

Model	Unstandardized Coefficients		T
	B	Std. Error	
(Constant)	284.482	78.146	3.640
EicherPriceLAG1	1.095	.062	17.516
EicherQuantityLAG1	-.001	.000	-3.469
HML PriceLAG10	-18.885	5.476	-3.448
HeroQuantityLAG7	-6.22E-05	.000	-3.416
MarutiQuantityLAG3	1.952E-05	.000	2.154
HML QuantityLAG2	2.450E-05	.000	2.154
HeroQuantityLAG8	4.852E-05	.000	2.477
EicherPriceLAG2	-.124	.061	-2.041
EicherQuantityLN	-3.523	4.430	-.795

Table IV indicates, in case of Eicher share prices in NSE, 8 explanatory automotive sector variables indicate significant causality.

TABLE V
REGRESSION ANALYSIS OF ***HERO***

Model	Unstandardized Coefficients		t
	B	Std. Error	
(Constant)	4.542265	47.3767	0.095876
HeroPriceLAG1	0.967113	0.013003	74.37759
HeroQuantityLAG1	-2.6E-05	8.18E-06	-3.23049
MarutiQuantityLAG6	1.28E-05	4.05E-06	3.151387
ALQuantityLAG7	1.04E-06	4.02E-07	2.572624
ALQuantityLAG4	-1.8E-06	5.01E-07	-3.53995
ALQuantityLAG3	1.58E-06	5.01E-07	3.151105
EicherQuantityLAG9	-0.00018	7.54E-05	-2.39466
BajajQuantityLAG6	-1.5E-05	7.88E-06	-1.89358
HeroQuantityLN	4.698555	3.507867	1.339434

In the case of Hero share prices in NSE, 8 explanatory automotive sector variables indicate significant causality

TABLE VI
REGRESSION ANALYSIS OF ***HINDUSTAN MOTORS***

Model	Unstandardized Coefficients		t
	B	Std. Error	
(Constant)	-0.42045	0.394217	-1.06655
HML PriceLAG1	0.758328	0.038465	19.71469
HML QuantityLAG1	-2.5E-07	3.36E-08	-7.32677
HML Quantity	2.01E-07	4.77E-08	4.21538
AL ClosePrice	0.163329	0.028187	5.794551
BajajPriceLAG4	0.001215	0.000276	4.403972
ALPriceLAG1	-0.1015	0.029174	-3.47917
BajajQuantityLAG2	-2E-07	4.89E-08	-4.16012
HeroQuantityLAG2	2.32E-07	5.01E-08	4.628412
BajajPriceLAG2	-0.00154	0.000341	-4.50408
BajajClosePrice	0.000816	0.000285	2.866716
BajajQuantityLAG9	1.52E-07	4.88E-08	3.118786
HMMPriceLAG4	0.075904	0.034411	2.205825
EicherQuantityLAG10	-9.6E-07	4.59E-07	-2.09809
ALQuantityLAG6	-6.8E-09	2.91E-09	-2.34206
TataQuantityLAG3	-3E-08	1.06E-08	-2.8619
MahindraPriceLAG9	-0.00096	0.000307	-3.13315
Hero ClosePrice	0.000296	0.000136	2.169438
EicherQuantityLAG5	9.65E-07	4.53E-07	2.12865
HMMQuantityLN	0.006223	0.02494	0.24953

In the case of Hindustan Motor share prices in NSE, 18 explanatory automotive sector variables indicate significant causality.

TABLE VII
REGRESSION ANALYSIS OF ***MAHINDRA***

Model	Unstandardized Coefficients		t
	B	Std. Error	
(Constant)	74.04359	34.50353	2.145972
MahindraPriceLAG1	0.818046	0.034337	23.82434
TataQuantityLAG8	2.57E-06	7.67E-07	3.356628
MarutiClosePrice	0.257828	0.033128	7.782677
MarutiPriceLAG1	-0.21065	0.036314	-5.8008
ALQuantityLAG4	-4.2E-07	2.19E-07	-1.93381
HeroQuantityLAG10	1.15E-05	3.89E-06	2.946424
MahindraQuantityLAG1	-4.6E-06	1.54E-06	-2.97954
MarutiQuantityLAG1	5.54E-06	2.07E-06	2.673347
MahindraPriceLAG9	0.104038	0.034152	3.046323
MarutiPriceLAG8	-0.03482	0.014165	-2.45844
ALQuantityLAG3	-1.3E-07	2.19E-07	-0.59516
MahindraQuantityLN	-1.86221	1.827656	-1.01891

In the case of Mahindra share prices in NSE, 11 explanatory automotive sector variables indicate significant causality.

TABLE VIII
REGRESSION ANALYSIS OF ***MARUTI***

Model	Unstandardized Coefficients		t
	B	Std. Error	
(Constant)	-140.755	58.76878	-2.39506
MarutiPriceLAG1	0.894518	0.020282	44.10304
MahindraClosePrice	0.584514	0.106392	5.493964
MahindraPriceLAG1	-0.34325	0.111102	-3.08949
MahindraQuantityLAG1	1.23E-06	2.57E-06	0.478714
Bajaj ClosePrice	0.259759	0.053126	4.889488
BajajPriceLAG1	-0.24048	0.054598	-4.40447
MahindraQuantityLAG10	-5.4E-06	2.56E-06	-2.12278
Bajaj Quantity	-1.3E-05	6.46E-06	-1.95602
MarutiQuantityLN	4.116857	3.147358	1.308036

In the case of Maruti share prices in NSE, 8 explanatory automotive sector variables indicate significant causality as shown in Table VIII.

TABLE IX
REGRESSION ANALYSIS OF ***TATA***

Model	Unstandardized Coefficients		t
	B	Std. Error	
(Constant)	-9.30845	7.126728	-1.30613
TataPriceLAG1	0.974064	0.011208	86.90671
Mahindra Quantity	-9.4E-07	3.5E-07	-2.68515
TataQuantityLAG5	-5.6E-07	1.78E-07	-3.16986
HML QuantityLAG4	-1.4E-06	5.27E-07	-2.63804
MahindraClosePrice	0.082345	0.013034	6.31786
MahindraPriceLAG1	-0.07346	0.013118	-5.60003
Bajaj Quantity	-2.1E-06	8.41E-07	-2.49075
TataQuantityLN	0.634161	0.377296	1.680806

In the case of Tata Motors stocks, 7 explanatory variables of the automotive equity indicate causality. The goodness of fit for each stock price regression model is depicted by the R square value in Table X.

TABLE X
GOODNESS OF FIR FOR EACH STOCK PRICE REGRESSION MODEL

Parameter	R Square
Stock	
Ashok Leyland	.995
Bajaj	.958
Eicher	.989
Hero	.978
Hindustan Motors Limited	.981
Mahindra	.939
Maruti	.971
Tata	.976

Thus, from the above regression models, it can be clearly observed that majority of automotive stock closing price value have a high correlation with the first lag of price and quantity variable. The closing price, quantities and of other automotive stocks and their respective lags also affect the price of a specific automotive share significantly and thus should be taken into consideration while making automotive sector equity investment decisions.

B. Decision Tree Classifiers

In the second part of the analysis, the data was subjected to two decision tree classifiers namely J48 and Random Tree. The J48 decision trees representing the decision making process for buying or selling shares for the respective stocks is depicted in Figs. 1 to 8.

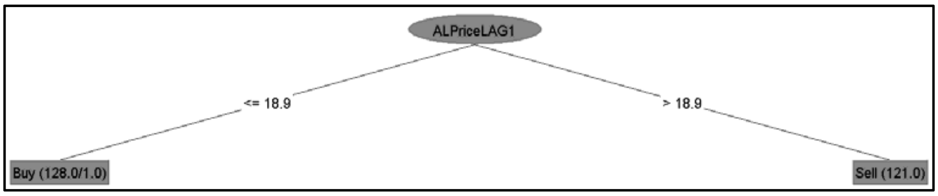

Fig. 1 J48 Tree for investment decisions in Ashok Leyland stocks

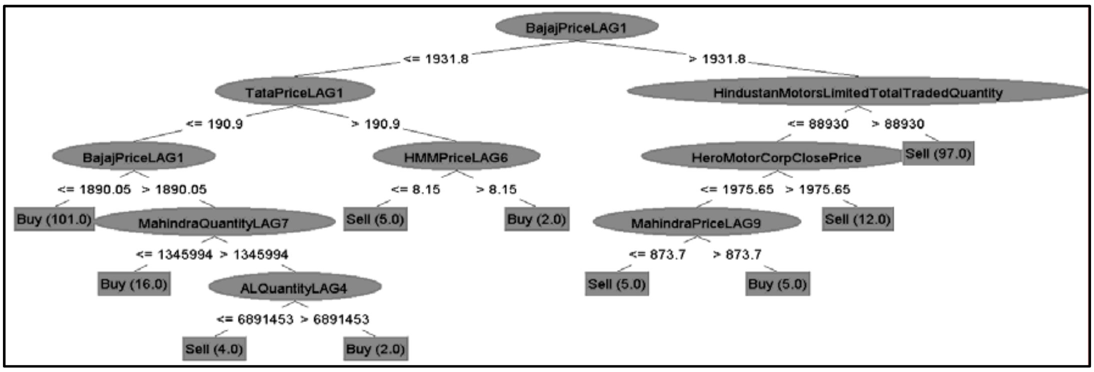

Fig. 2 J48 Tree for investment decisions in Bajaj stocks

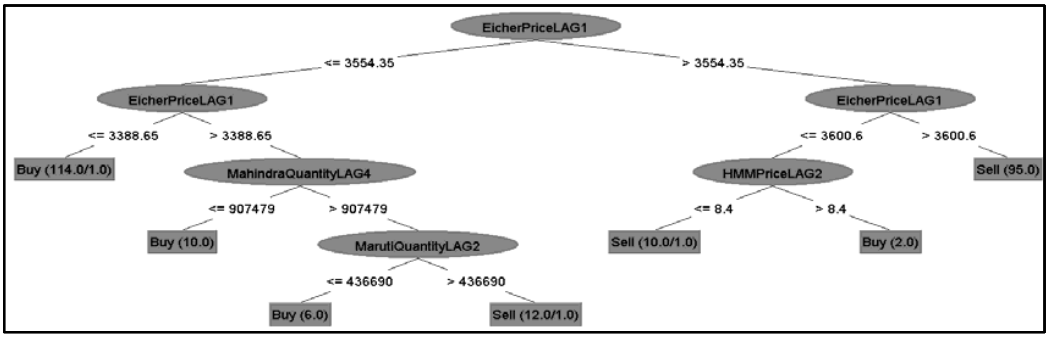

Fig. 3 J48 Tree for investment decisions in Eicher stocks

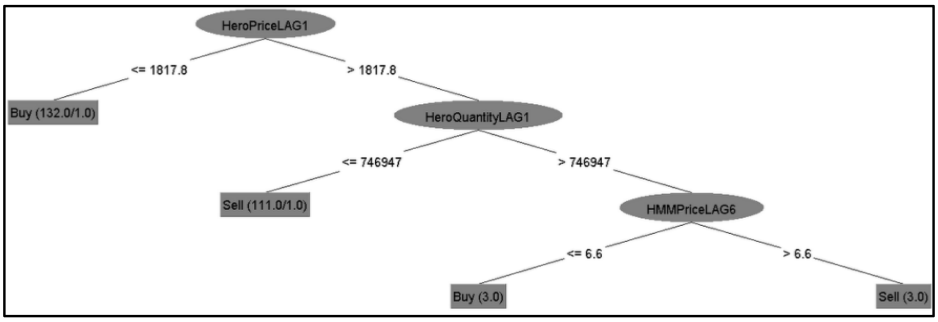

Fig. 4 J48 Tree for investment decisions in Hero stocks

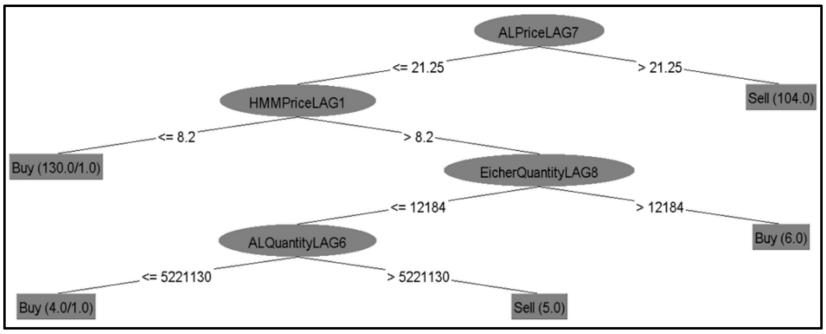

Fig. 5 J48 Tree for investment decisions in Hindustan Motors stocks

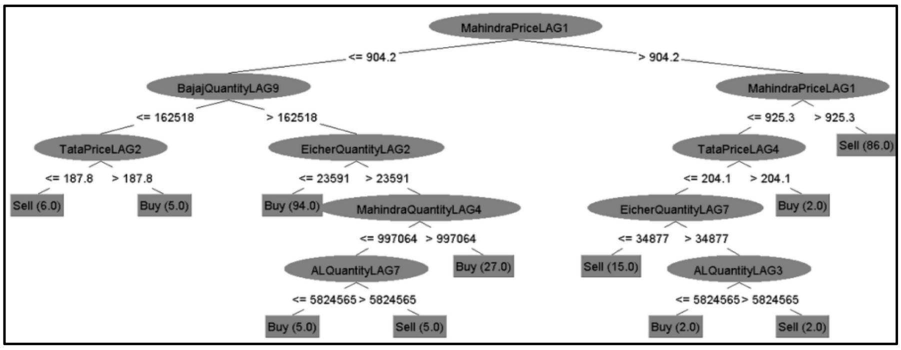

Fig. 6 J48 Tree for investment decisions in Mahindra stocks

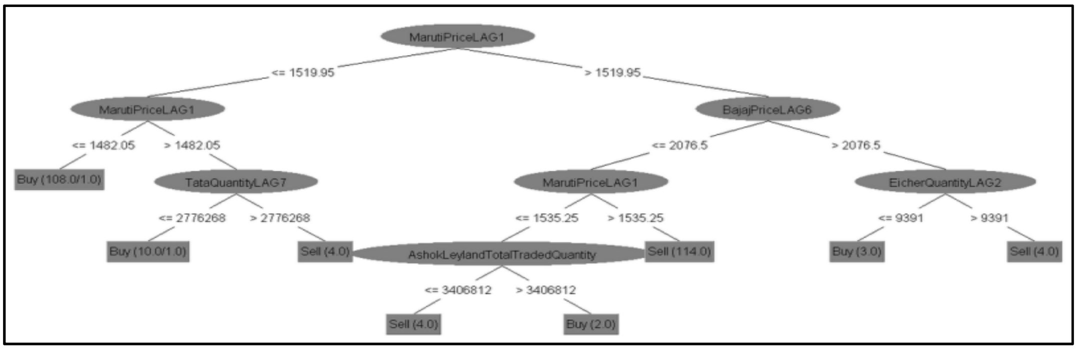

Fig. 7 J48 Tree for investment decisions in Maruti stocks

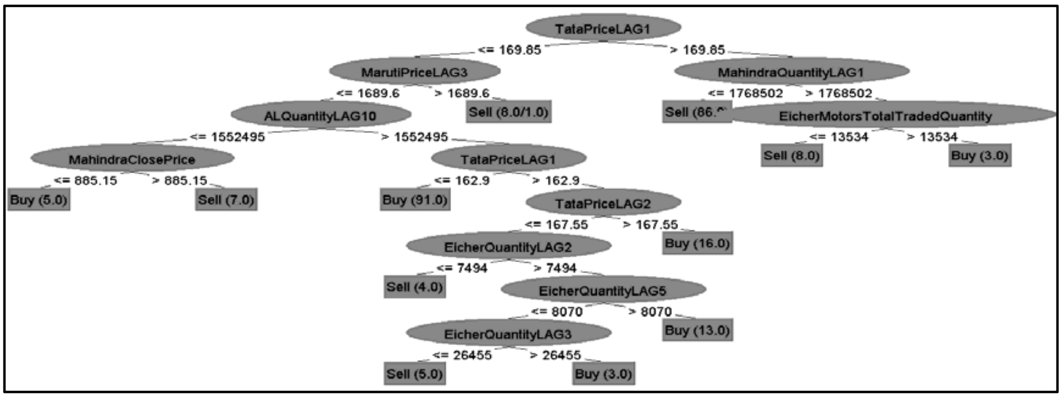

Fig. 8 J48 Tree for investment decisions in Tata stocks

The J48 tree for Bajaj stocks in Fig. 2 indicates that price is influenced not just by Bajaj quantity but first by Tata prices and by Hindustan quantity. In the case of Eicher stocks in Fig. 3, the Eicher price lags have a significant influence on the prices. In the case of the Hero motor stocks in Fig. 4, the prices are determined by its price and quantity lags only and not affected by any other variable. In the case of the Hindustan Motors stocks in Fig. 5, the Ashok Leyland prices have a significant influence on the prices of Hindustan Motors stocks. In Fig. 6, the lags of Mahindra stock prices have a significant influence on the prices of Mahindra stocks. The J48 tree for Maruti stocks in Fig. 7 indicate that the prices are influenced by the Bajaj prices and Tata Motors stock prices. In the case of Tata share prices in Fig. 8, the tree indicates that the share prices of Tata are influenced by Mahindra and Maruti.

The random trees generated for the data were too big to be included in the article and thus, the significant variables which appeared in the random tree and J48 tree have been listed in Table XI.

TABLE XI
SIGNIFICANT VARIABLES FOR TREE CLASSIFICATION

Stock	Tree	
	J48	**Random Tree**
Ashok Leyland	AL Price LAG1	Hero Price LAG8 AL Price LAG9 Mahindra Price LAG6 Hero Price LAG2
Bajaj	Bajaj Price Lag1 Tata Price LAG1 HML Quantity Bajaj Price LAG1 HML Price LAG6 Hero Close Price Mahindra Quantity LAG7 Mahindra Price LAG9 AL Quantity LAG4	Bajaj Price LAG 3 Maruti Quantity LAG 4 Tata Close Price Eicher Price LAG7 Hero Price LAG5 Maruti Price LAG4 Bajaj Price LAG4 HML Quantity LAG7 Mahindra Price LAG10 HML Quantity LAG8 AL Close Price Mahindra Price LAG5 Bajaj Price LAG9 AL Price LAG8 Eicher Price LAG6 AL Quantity LAG3
Eicher	Eicher Price LAG1 Mahindra Quantity LAG4 HML Price LAG2 Maruti Quantity LAG4 Maruti Quantity LAG2	AL Price LAG8 Eicher Price LAG8 Bajaj Price LAG1 Tata Price LAG8 Mahindra Price LAG8 AL Price LAG2 Maruti Price LAG4 Eicher Price LAG6 Bajaj Price LAG8 AL Price LAG1 Bajaj Price LAG6 Mahindra Price LAG6 AL Price LAG10
Hero	Hero Price LAG1 Hero Quantity LAG1 HML Price LAG6	Tata price LAG3 Ashok Leyland Close Price Hero Price LAG 2 Bajaj Quantity LAG10 Eicher Price LAG1 Hero Price LAG2 HML price LAG4 Tata Quantity LAG8 Eicher Price LAG4
Hindustan Motors	AL Price LAG7 HML Price LAG1 Eicher Quantity LAG8 AL Quantity LAG6	HML Price LAG9 AL Price LAG4 Hero Price LAG4 Hero Quantity LAG9 Tata Price LAG4 HML Quantity AL Close Price Bajaj Price LAG9 Maruti Price LAG7 Eicher Quantity LAG5 Eicher Price LAG8
Mahindra	Mahindra Price LAG1 Bajaj Quantity LAG9 Mahindra Price LAG1 Tata Price LAG2 Eicher Quantity LAG2 Tata Price LAG 4 Eicher Quantity LAG7 Mahindra Quantity LAG4 AL Quantity LAG7 Al Quantity LAG3	Mahindra Price LAG4 Eicher Price LAG9 Maruti Quantity LAG3 AL Price LAG6 Hero Price LAG9 Maruti Quantity LAG5 Eicher Quantity LAG1 HML Price LAG4 Mahindra Price LAG4 HML Quantity LAG10 Tata Price LAG1 Hero Price LAG5 HML Price LAG4 Tata Price LAG2 Bajaj Quantity LAG9 Mahindra Price LAG3 AL Quantity LAG7 Tata Price LAG6 AL Quantity LAG10 Bajaj Quantity LAG10 Eicher Quantity LAG5 AL Price LAG1
Maruti	Maruti Price LAG1 Bajaj Price LAG6 Tata Quantity LAG7 Eicher Quantity LAG2 AL Quantity	Tata Price LAG1 Hero Price LAG7 Eicher Price LAG8 AL Price Lag7 Hero Price LAG10 Mahindra Price LAG2 Mahindra Close Price Eicher price LAG8 Tata Price LAG7 AL Price LAG5 Hero Price LAG8 Tata Price LAG1 Bajaj Quantity LAG2 Maruti Price LAG5 Hero Quantity LAG3 Tata Price LAG2
Tata	Tata Price LAG1 Tata Price LAG2 Maruti Price LAG3 Mahindra Quantity LAG1 AL Quantity LAG10 Eicher Quantity Eicher Quantity LAG2 Eicher Quantity LAG3 Eicher Quantity LAG5	Tata Price LAG5 HML Price LAG2 Tata Price LAG1 AL Quantity LAG8 AL Quantity LAG4 Tata Price LAG9 Eicher Price LAG9 Hero Price LAG10 Hero price LAG9 Maruti Price LAG9 Bajaj Close Price HML Quantity LAG8 Maruti Price LAG4 Bajaj Price LAG10 HML Price LAG7 Bajaj Quantity LAG2 AL Price LAG3 Hero Quantity LAG4 Bajaj Price LAG9 Bajaj Quantity LAG4 Eicher Quantity LAG2 AL Price LAG2

The respective classification accuracies obtained for each respective stock using a ten-fold cross validation after training the classifier corresponding to each algorithm are presented in Table XII.

TABLE XII
CLASSIFICATION ACCURACIES OF DECISION TREES FOR RESPECTIVE STOCKS

Parameter Stock	Fit of Trees	
	J 48	**Random**
Ashok Leyland	98.39%	98.80%
Bajaj	88.76%	84.34%
Eicher	90.36%	88.76%
Hero	96.79%	95.58%
Hindustan Motors Limited	94.38%	93.57%
Mahindra	88.76%	85.94%
Maruti	90.76%	88.76%
Tata	81.12%	77.91%

It is clearly evident that in all the stocks, the classification accuracy is mostly higher for the J48 algorithm and almost equal for both the classifiers in case of one stock. Thus, we can infer that J48 is an efficient classification and decision making technique for buy sell decisions in the automotive sector stocks.

V. CONCLUSION

From the above study, the interdependence of stock prices in the automotive sector on each other is clearly evident. The closing price of a stock also depends on the lags of its own price and quantity as well as, the price and quantity of other stocks. The buying and selling decisions involved in an automotive equity portfolio take in to account values of these variables as depicted in the decision tree. Each of the automotive stock prices has different influencers in the industry. It is important that the intra-sector factors have a very significant role in the price determination

ACKNOWLEDGMENT

The authors of this paper would like to acknowledge the contributions of Kaavya T R, Vinod K, Siva Rama Krishna, currently pursuing post-graduate program at VIT Business School for helping us in compiling the stock market data for analysis.

REFERENCES

[1] Jar-Long Wang, Shu-Hui Chan, “Stock market trading rule discovery using two-layer bias decision tree”, *Expert Systems with Applications*, vol. 30, pp 605-611, 2006 .

[2] Chih-Fong Tsai and Yu-Chieh Hsiao, “Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches”,*Decision Support Systems*, vol. 50, pp 258-269, 2010.

[3] TomerGeva and Jacob Zahavi, “Emperical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news”, *Decision Support Systems*, vol. 57, pp 212-223, 2014.

[4] Man Hong Wong, “Decision Support Investment models based on clustered scenario trees”, 2014, *European journal of Operations Research*, article in press.

[5] Muh-Cherng Wu, Sheng-Yu Lin, Chia-Hsin Lin, “An effective application of decision tree to stock trading”,*Expert Systems with Applications*, vol. 31, pp. 270–274, 2006.

[6] Robert K. Lai , Chin-Yuan Fan , Wei-Hsiu Huang , Pei-Chann Chang, ”Evolving and clustering fuzzy decision tree for financial time series data forecasting”, *Expert Systems with Applications*, vol. 36, pp. 3761–3773, 2009.

[7] Suh Son, KyongJooOha, Tae Yoon Kim and Dong Ha Kim,“An early warning system for global institutional investors at emerging stock markets based on machine learning forecasting”, *Expert Systems with Applications*, vol. 36, pp. 4951–4957, 2009.

[8] Pei-Chann Chang, Chin-Yuan Fan and Jun-Lin Lin,“Trend discovery in financial time series data using a case based fuzzy decision tree”, *Expert Systems with Applications*, vol. 38, pp. 6070–6080, 2011.

[9] Wen-Shiung Lee, Alex YiHou Huang , Yong-Yang Chang and Chiao, Ming Cheng,“Analysis of decision making factors for equity investment by DEMATEL and Analytic Network Process”,*Expert Systems with Applications*, vol. 38, pp. 8375–8383, 2011.

[10] David Diaz, BabisTheodoulidis, Pedro Sampaio,“Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices”,*Expert Systems with Applications*, vol. 38, pp. 12757–12771, 2011.

[11] Tsung-Sheng Chang, “A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction”, *Expert Systems with Applications*, vol. 38, pp. 14846–14851, 2011.

[12] WangrenQiu, Xiaodong Liu, LidongWang,“Forecasting shanghai composite index based on fuzzy time series and improved C-fuzzy decision trees”,*Expert Systems with Applications*, vol. 39, 7680–7689, 2012.

[13] Shu-Hsien Liao, Shan-Yuan Chou,“Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio”,*Expert Systems with Applications*, vol. 40, pp. 1542–1554, 2013.

[14] Chih-Fong Tsaia, Yuah-ChiaoLinc, David C. Yenb, Yan-Min Chenc, “Predicting stock returns by classifier ensembles”,*Applied Soft Computing*, vol. 11,pp. 2452–2459.

[15] PreetiParanjape, Voditel, UmeshDeshpande,“A stock market portfolio recommender system based on association rule mining”,*Applied Soft Computing*, vol. 13, pp. 1055–1063.

[16] Agnes Virlics, “Investment Decision Making and Risk”,*Procedia Economics and Finance (Science Direct)*, vol. 6, pp. 169 – 177.