

Regression Analysis Project

By Vardhan Babburi, Prasanna Muppidi

INTRODUCTION

RESEARCH BACKGROUND

RESEARCH MOTIVATION

- Portfolio Optimization is very important. Advisable to diversify investments across not just equity but also commodity markets.
- Commodity markets can be used to gauge Dow Jones Price Movement?
- Technical analysis is a way to forecast the markets. This helps traders and investors understand future price movement across variety of time frames.
- There are a whole lot of technical indicators.
- Which are the ones that should be a “must use” for understanding and predicting Dow Jones?

RESEARCH QUESTIONS

- **Q1: Predicting Dow Jones with the help of Crude and Gold previous day prices.**
- **Q2: Which are the most important technical indicators to predict market bullish or bearish?**

PROBLEM WE ARE SOLVING:

- Help predict Dow Jones market sentiment and Price with available data like:
 1. Crude Oil previous Day Price
 2. Gold Previous Day Price
 3. Technical Indicators on previous Close Prices (lag) of Dow Jones

METHODOLOGY

DATA COLLECTION

- Yahoo Finance
 - Investing.com
-
- Data from 2017 to April 2022 - Q1
 - Data from 2010 to April 2022 - Q2

DATA PREPARATION

Question 1

- SAS Data Step
- CSV files loaded for Gold, Crude and Dow
- Merged using Date
- Using Lag function to calculate previous day close prices

Question 2

- SAS Data Step
- Technical Indicators CSV file loaded

ANALYSIS METHOD - CART

- Descriptive and correlation analysis conducted.
- Applied Classification and Regression Trees

Question 1: Dependent Variable - Dow Jones Close Price today, Independent Variable: Previous Day Close price of Gold and Crude

Question 2: Dependent Variable: Dow Jones Up/Down, Independent Variables: RSI, ADX, SMA and LMA

CLASSIFICATION & REGRESSION TREES

- Regression and classification trees are methods for analyzing how a dependent variable is correlated with independent variables. PROC HPSPLIT and PROC DTREE can both create decision trees that look similar.
- When we want to explore the relationship of variables and outcome, that is the effect of variables on the outcome, PROC HPSPLIT is a useful tool.
- Decision trees are a machine learning technique for making predictors; they are built by repeatedly splitting data into smaller and smaller clusters. The trees are trained by passing data down from a root node to leaves. The data is repeatedly split based on predictor variables therefore the sub nodes would be more homogenous.

DATA ANALYSIS & FINDINGS

Variable	Mean	Std Dev	Minimum	Maximum	N
close_dow	25304.87	2806.79	18591.93	32778.64	1055
close_crude	53.2165309	11.5724515	-37.6300011	76.4100037	1055
close_gold	1441.20	237.2770215	1160.40	2051.50	1054

3 Variables: close_dow close_gold close_crude

Pearson Correlation Coefficients Number of Observations			
	close_dow	close_gold	close_crude
close_dow	1.00000 1055	0.63989 1054	0.15356 1055
close_gold	0.63989 1054	1.00000 1054	-0.54520 1054
close_crude	0.15356 1055	-0.54520 1054	1.00000 1055

QUESTION 1:

Dow Jones Close ~

- Previous Day Crude Price,
- Previous Day Gold Price).

Gold Price more important factor.

Variable Importance			
Variable	Training		Count
	Relative	Importance	
close_pd_gold	1.0000	62123.9	5
close_pd_crude	0.8264	51338.4	2

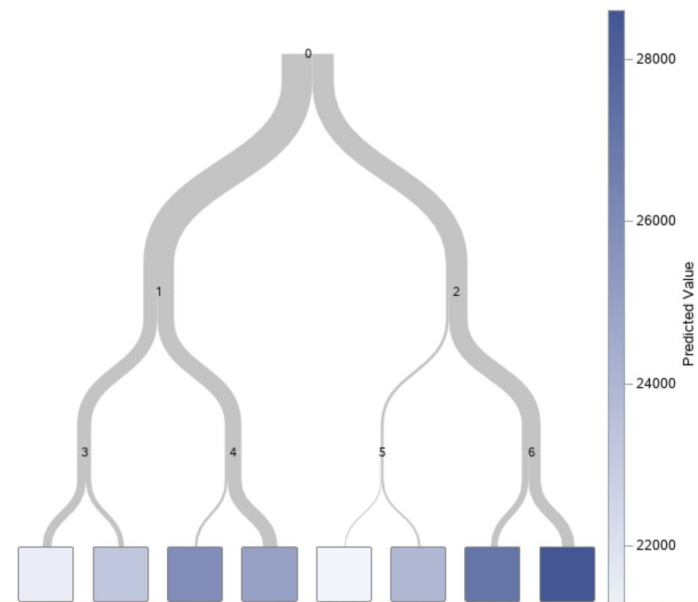
Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	Engine	Role	Path
WORK.DUMMY	V9	Input	On Client
WORK.RESULT1	V9	Output	On Client

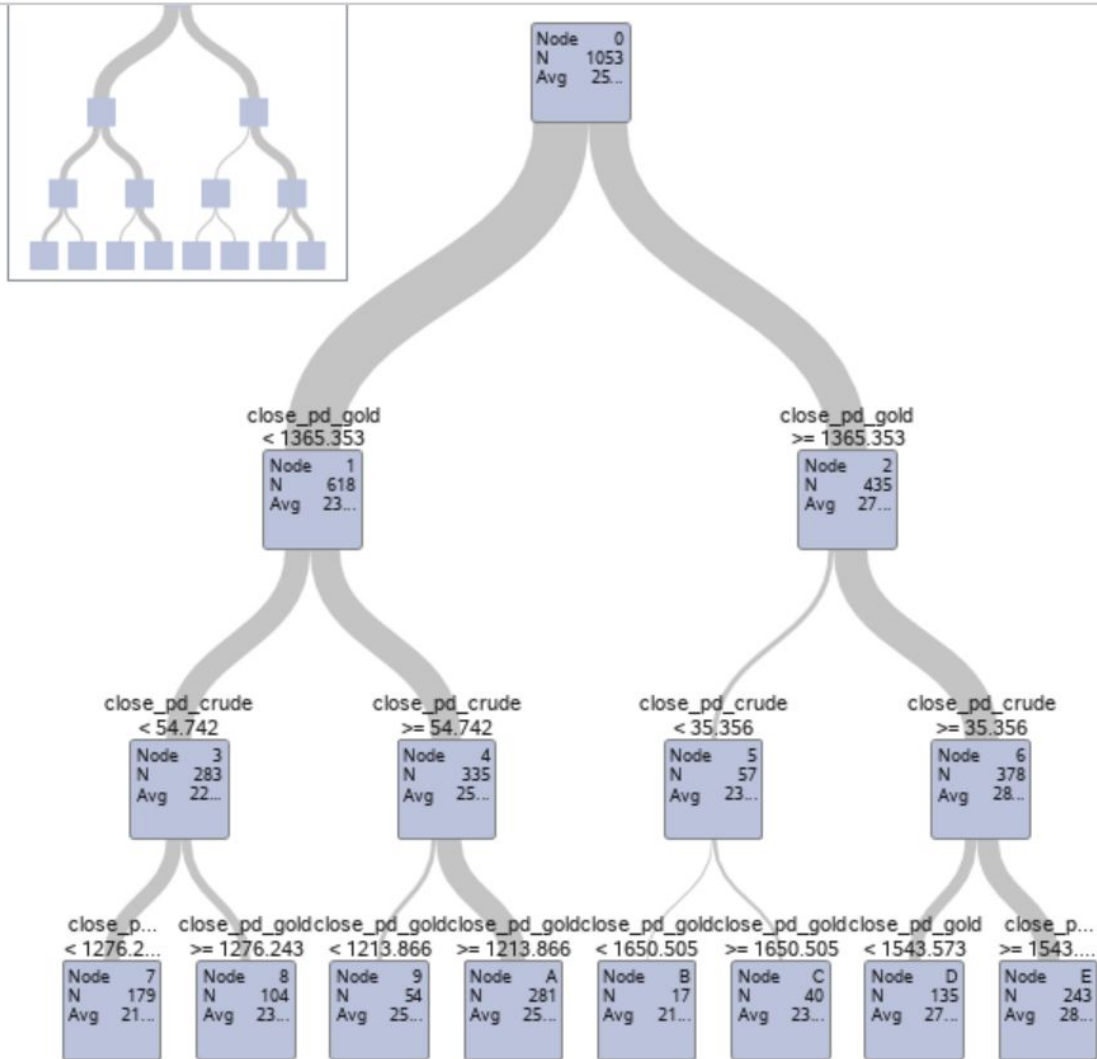
Model Information	
Split Criterion Used	Variance
Pruning Method	Entropy
Subtree Evaluation Criterion	Full Tree
Number of Branches	2
Maximum Tree Depth Requested	3
Maximum Tree Depth Achieved	3
Tree Depth	3
Number of Leaves Before Pruning	8
Number of Leaves After Pruning	8

Number of Observations Read	1055
Number of Observations Used	1053

Regression Tree for close_dow



If Previous Gold Price greater than 1365 and Previous Crude Price is greater than 35.36 and previous gold price greater than 1543 then the Dow Jones has got very high probability to close towards higher prices (blue spectrum).



QUESTION 2:

Most Important indicators RSI and ADX.

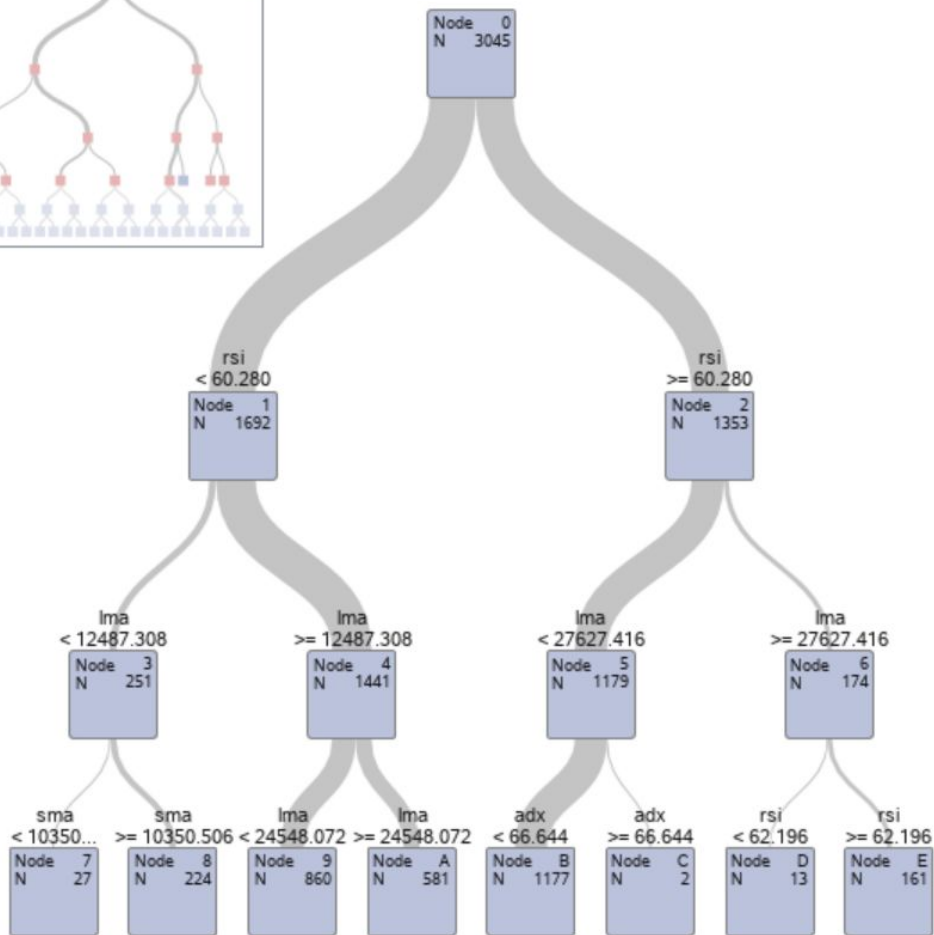
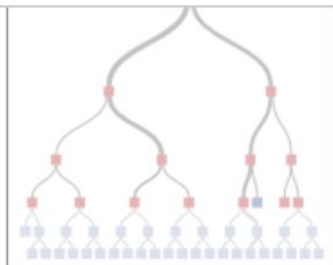
Dependent Variable: Up/Down

Independent Variables: RSI, ADX, SMA, LMA

Variable Importance			
Variable	Training		Count
	Relative	Importance	
rsi	1.0000	5.4999	9
adx	0.6860	3.7729	6
lma	0.6497	3.5731	7
sma	0.5009	2.7551	2

Model-Based Confusion Matrix			
Actual	Predicted		Error Rate
	Down	Up	
Down	34	577	0.9444
Up	24	2410	0.0099

Model-Based Fit Statistics for Selected Tree								
N Leaves	ASE	Mis-class	Sensitivity	Specificity	Entropy	Gini	RSS	AUC
25	0.1497	0.1974	0.0556	0.9901	0.6755	0.2995	912.0	0.6659



Return 1=Down 2=Up

If RSI is greater than 60.2, and LMA is greater than 27627, and RSI is greater than 62.1, then Market is Up!

CONCLUSION

- Investors and Traders participating in the U.S. equity markets can take help of Previous Day Gold and Crude prices to forecast next day close price using our Model 1.
- Investors and Traders can gauge price movement and sentiment of Dow Jones by making use of RSI and ADX.

THANK YOU!