STOCK VALUE PREDICTION

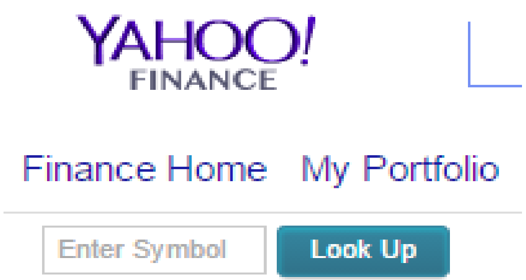
- Siddhant Sanjeev (15BM6JP45)
- Loveleen Kaur (15BM6JP21)
- Ankit Sonthalia (15BM6JP06)

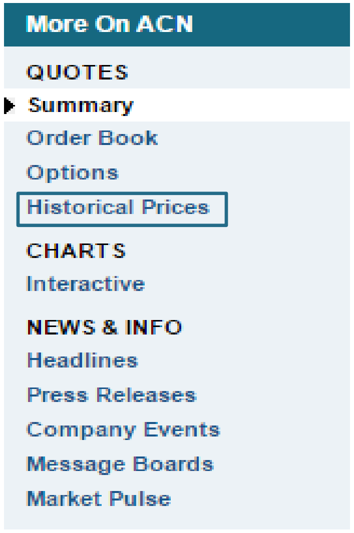
All characters and events depicted in this project are entirely modeled. Any similarity to actual events or stock price movements is purely awesome.

OUTLINE

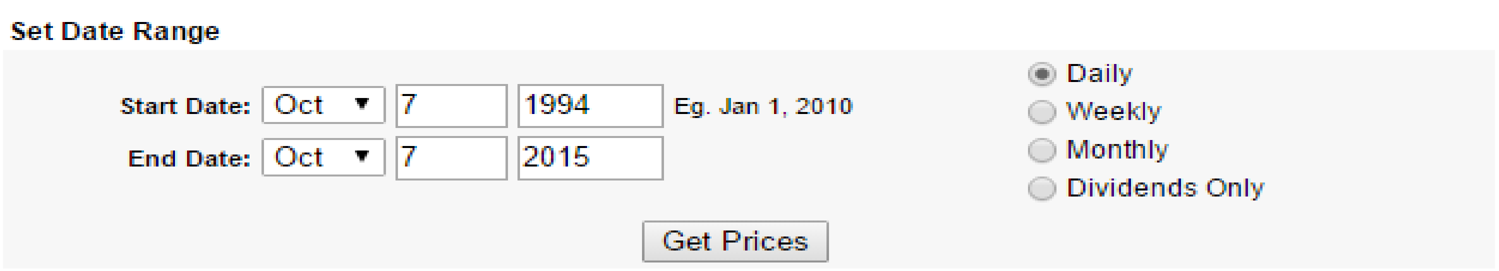
BAISI1/2	• Implementing probabilistic and statistical models
BAISI3	• Making inferences from models
BAISI4	• Computing integration of models
BAISI5	• Creating database of predictions

STOCKS AND MARKET DATA





http://finance.yahoo.com/q?s=DATA



POLYNOMIAL MODEL

Degree=5

Linear Model

```
input1=(1:length);
input2=input1^2;
input3=input1^3;
input4=input1^4;
input5=input1^5;

tsstock=ts(st);
lmstock=tslm(tsstock~input1+

input2+input3+

input4+input5)
```

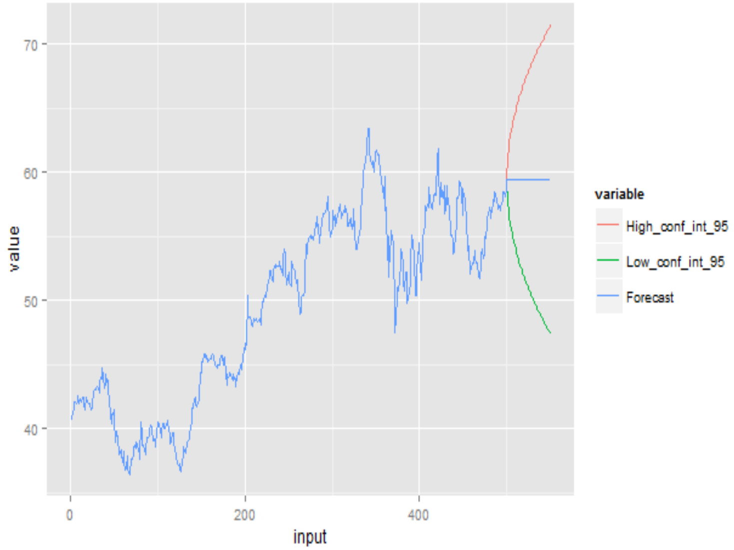
R Code Snippet

Fits the n^{th} degree polynomial on data points

Here n is chosen to be *5*

60
50
40
30
20
value
0
500
1000
input
variable
Polynomial_Model
Stock_Value

SES MODEL



Forecast with 95% C.I.

```
tsstock=ts(st);
 sesstock=ses(tsstock)

sesstock$fitted
#Forecasts

sesstock$upper
#High Confidence Interval

sesstock$lower
#Low Confidence Interval
```

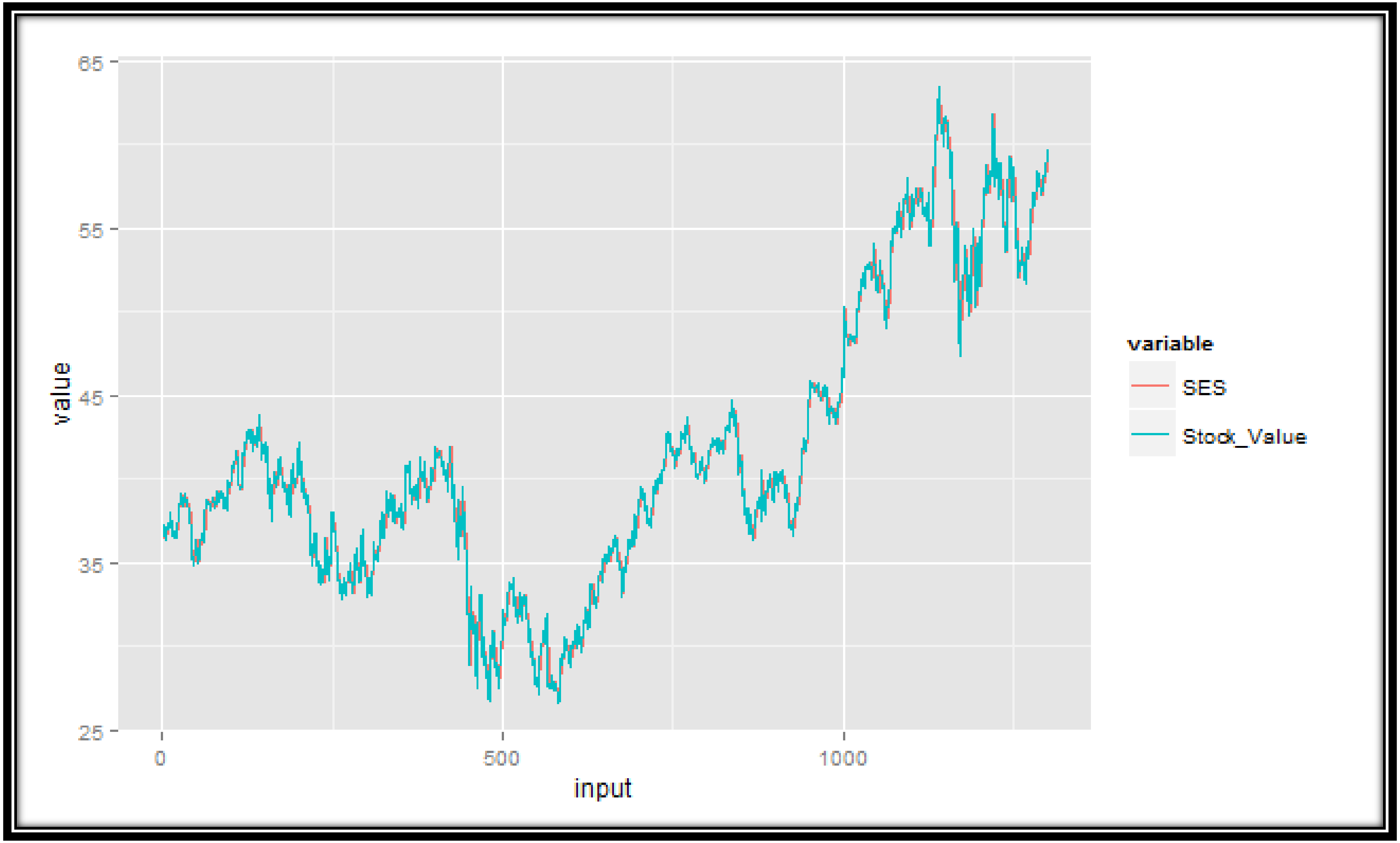
R Code Snippet

$\hat{y}_{T+1|T} = \alpha Y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2}$ .....

$\hat{y}_{T+1|T} = \alpha Y_T + (1-\alpha)\, \hat{y}_{T|T-1}$

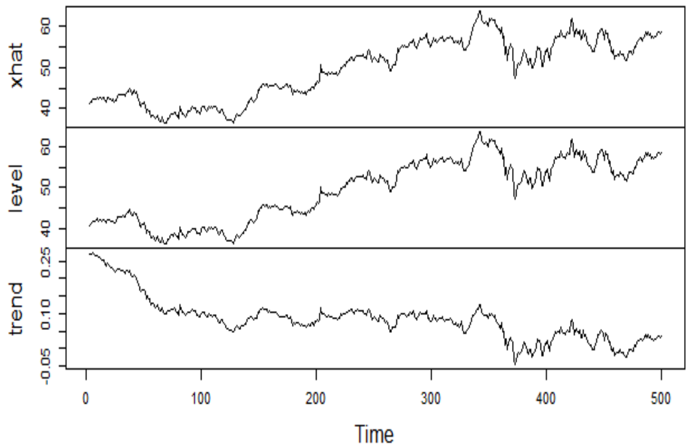
α= Smoothing Parameter

Weights decrease exponentially



65
55
45
35
25
value
0
500
1000
input
variable
SES
Stock_Value

HOLT-WINTERS MODEL



Holt-Winters Model

```
tsstock=ts(st);

hwstock=HoltWinters(tsstock,
                    gamma=FALSE)

forecast = predict(hwstock,
                   n.ahead = 100,

prediction.interval=T,
                   level = 0.95)

Holt_Wintrs_Fit=hwstock$fitted
```

R Code Snippet

Breaks data into level(a), trend(b) and seasonality(s)

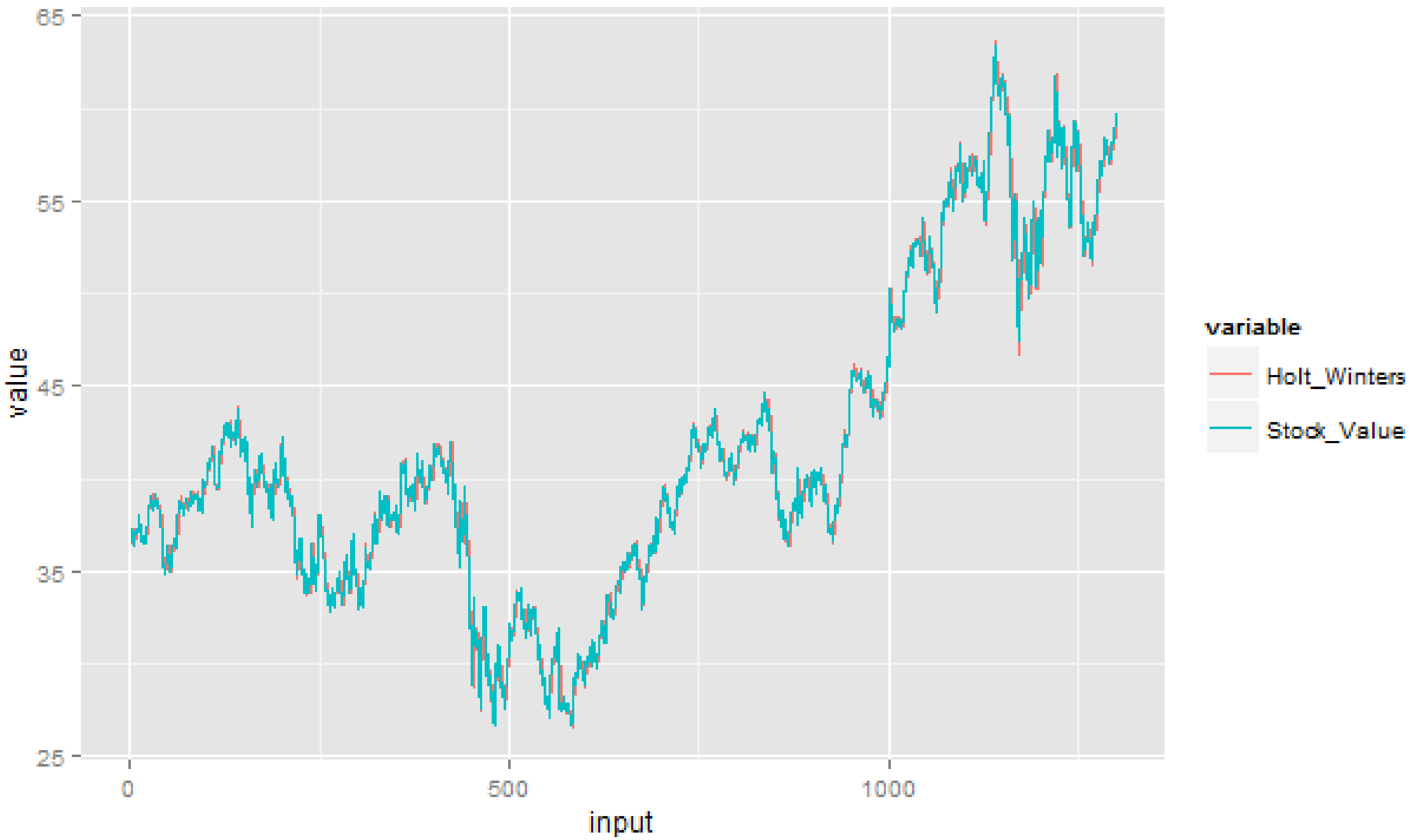
α=Smoothing Parameter | β=Exponential Smoothing | γ=Seasonal component

$$Y[t+h] = a[t] + h * b[t] + s[t - p + 1 + (h - 1) \bmod p]$$

$$a[t] = \alpha (Y[t] - s[t-p]) + (1-\alpha) (a[t-1] + b[t-1])$$

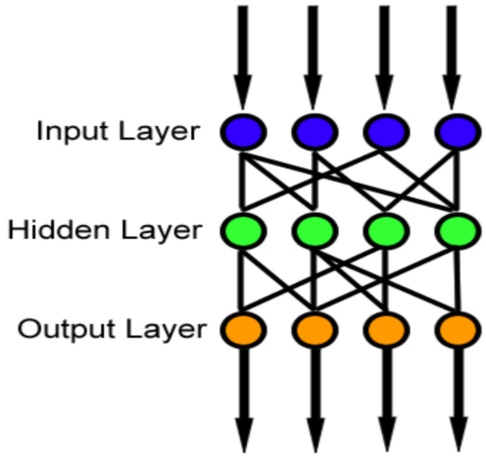
$$b[t] = \beta (a[t] - a[t-1]) + (1-\beta) b[t-1]$$

$$s[t] = \gamma (Y[t] - a[t]) + (1-\gamma) s[t-p]$$



65
55
45
35
25
value
0
500
1000
input
variable
Holt_Winters
Stock_Value

FEED FORWARD MODEL



Feed Forward Model

```
tsstock=ts(st);

 nnstock=nnetar(tsstock);
 predict=forecast(nnstock);

predict$mean[1]
#Predicts the value
```

R Code Snippet

Artificial Neural Network
No direction cycle
Flow of info in 1 direction : Forward

65
55
45
35
25
value
0
500
1000
input
variable
Neural_Network
Stock_Value

ARIMA MODEL

ar1	-1.59241	ar10	-0.894651
ar2	-2.392885	ar11	-0.611519
ar3	-2.312248	ar12	-0.393537
ar4	-2.180835	ar13	-0.213871
ar5	-2.162685	ar14	-0.086129
ar6	-1.891472	ar15	-0.00816
ar7	-1.684254	ma1	-0.336171
ar8	-1.394506	ma2	0.254239
ar9	-1.157487	ma3	-0.958154

ARIMA Coefficients

```
tsstock=ts(st);

arstock=arima(tsstock,
                 order=c(15,3,3),

method=c("CSS"));

pred=forecast(arstock);
pred$mean[1]
```

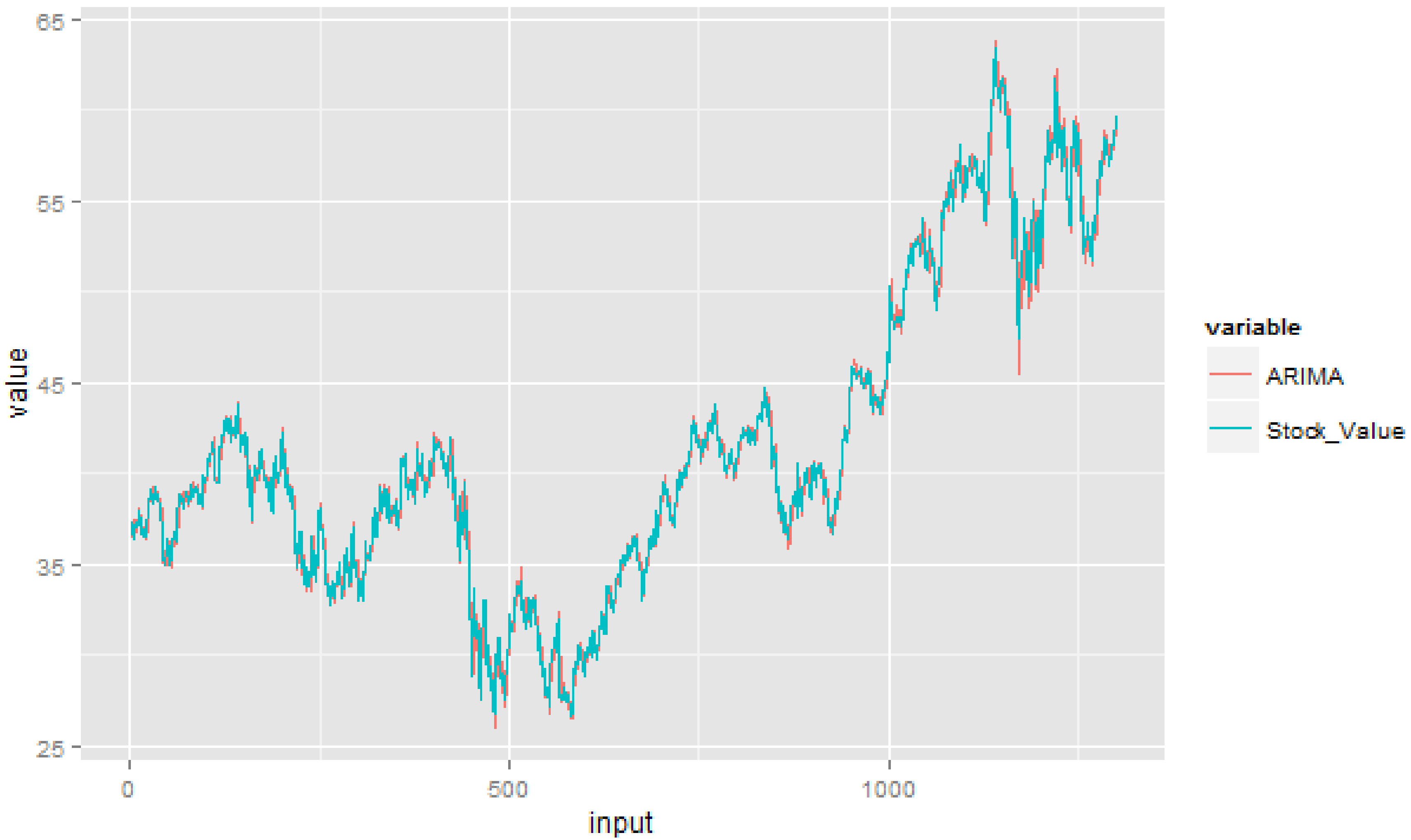
R Code Snippet

Autoregressive Integrated Moving Average(p,d,q)

p=AR order | d=degree of differencing | q=MA order

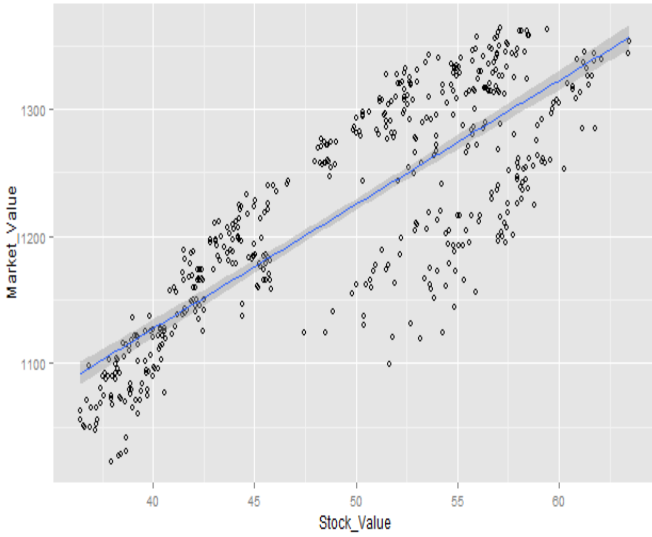
ARMA: $X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \ldots + \theta_q W_{t-q}$

Difference between d^{th} order ARMA is ARIMA



65
55
45
35
25
value
0
500
1000
input
variable
ARIMA
Stock_Value

MARKET-STOCK CORRELATION



Regression Line

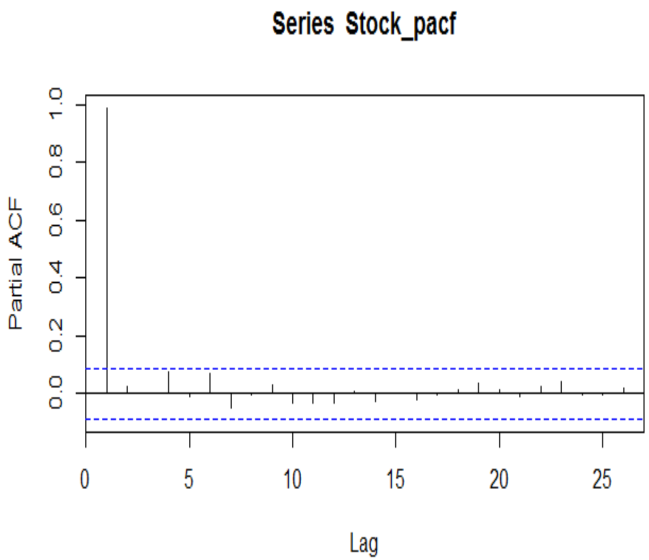
```
mar=as.numeric(k+1);
for(j in (i-k):i-1)
{
   mar[j-(i-k)+1]=market[j];
}

corstock=lm(stock~mar)
corstock=ts(corstock)

pred=forecast(corstock$fitted.
                         values)
```

R Code Snippet

The scatter plot between market and stock value is plotted
Linear regression is calculated on it

65
55
45
35
25
value
0
500
1000
input
variable
Market_Prediction
Stock_Value

STOCK AUTOCORRELATION

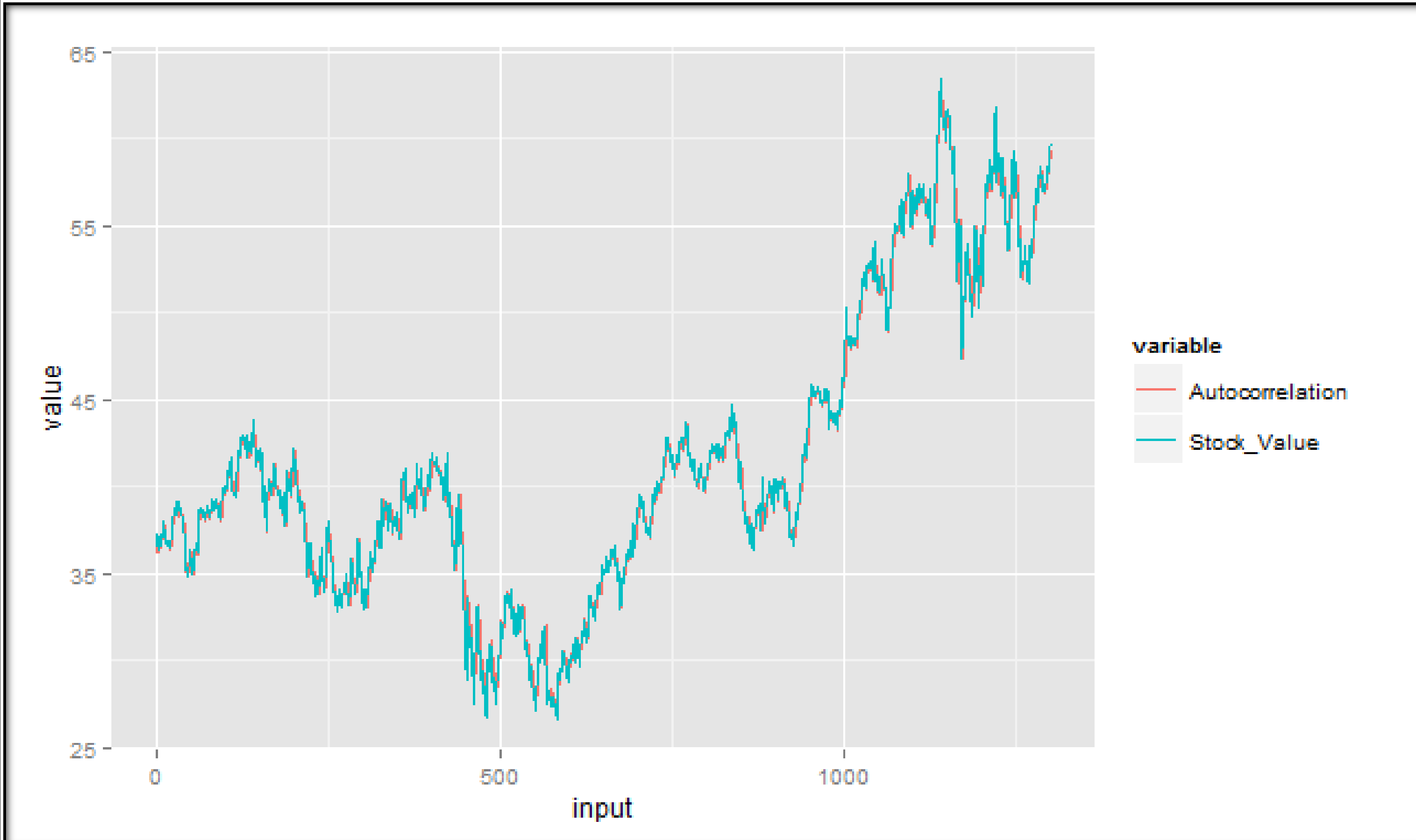


Autocorrelation Values

```
for(j in 2:k)
  temp[j-1]=st[j];

value[p]=cor(temp,st)
p=p+1;

corstock=ts(lm(st~temp))

pred=forecast(corstock$fitted.
                 values)
```

R Code Snippet

The scatter plot between stock value at time t_n and t_{n-1} is plotted
Linear regression is calculated on it



65
55
45
35
25
value
0
500
1000
input
variable
Autocorrelation
Stock_Value

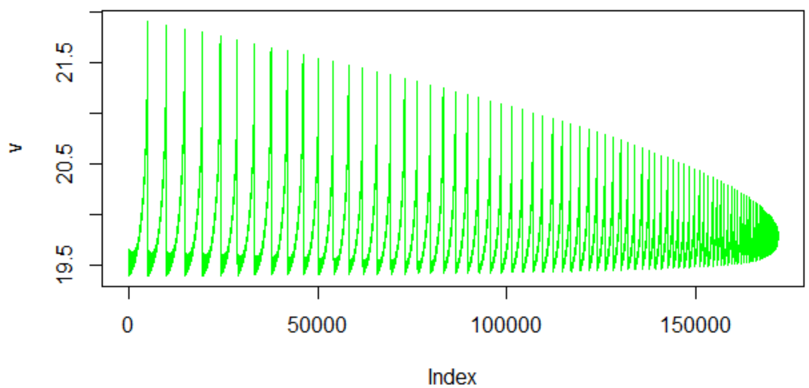
CORRELATION TABLE

- **Correlation Table** provides a linear correlation between the predicted value and real stock values.
- Higher the Autocorrelation better the estimated values.

Model No	Model Name	Correlation values
1	Polynomial Model	0.9462837
2	Holt-Winters Model	0.9956402
3	Feed Forward Model	0.9587529
4	ARIMA Model	0.9951617
5	SES Model	0.9957231
6	Market-Stock Correlation	0.9273835
7	Stock Autocorrelation	0.9957109

COMPUTING INTEGRATION OF MODELS

- Weighted average of the predicted models



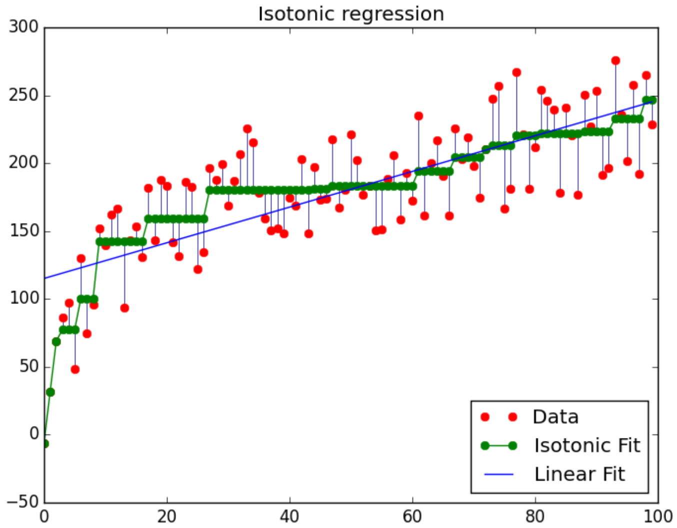

```c
    }

    for(w1=0;w1<=1;w1=w1+0.01)
    {
        for(w2=0;w2<=1;w2=w2+0.01)
        {
            if(w1+w2>1)
            break;

            for(w3=0;w3<=1;w3=w3+0.01)
            {
                if(w1+w2+w3>1)
                break;

                w4=1-w1-w2-w3;

                sum=0;
                for(i=1;i<=1300;i++)
                {
                    x=w1*e1[i]+w2*e2[i]+w3*e3[i]+w4*e4[i];
                    sum=sum+((x-r[i])*(x-r[i]))/r[i];

                }

                printf("%lf %lf %lf %lf %lf\n",w1,w2,w3,w4,sum);
            }
        }
    }

```

C Code Snippet

Applying different set of weights to the selected models.
Calculation of Chi-square value and selecting appropriate weight for the prediction

value
90
80
70
60
0
200
400
600
input
variable
x
y

➢ISOTONIC REGRESSION



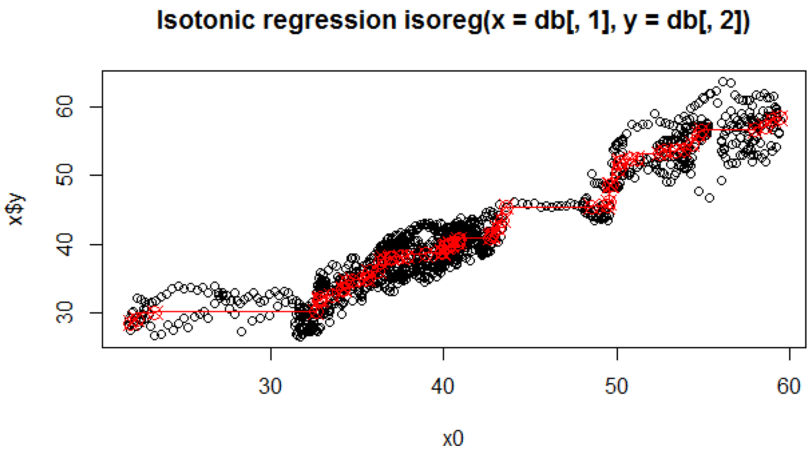

```r
x1=isoreg(db[,1],db[,2])
x2=isoreg(db[,3],db[,4])
x3=isoreg(db[,5],db[,6])
x4=db[,7]
y1=isoreg(x1$y,x2$y)
y2=isoreg(x3$y,x4)
z=isoreg(y1$y,y2$y)
```

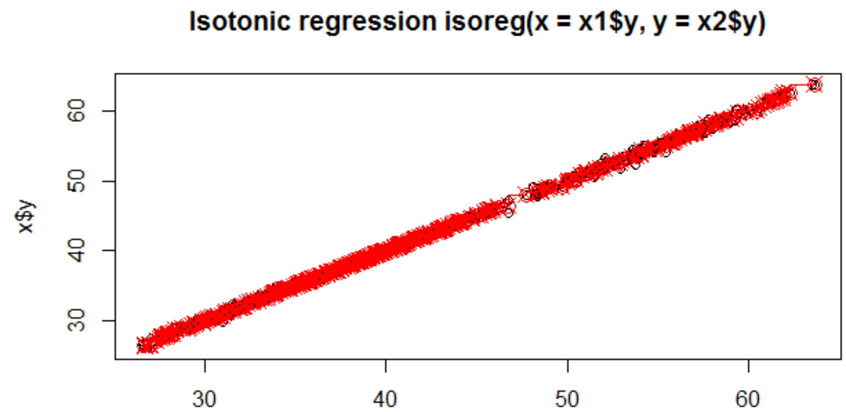
R Code Snippet

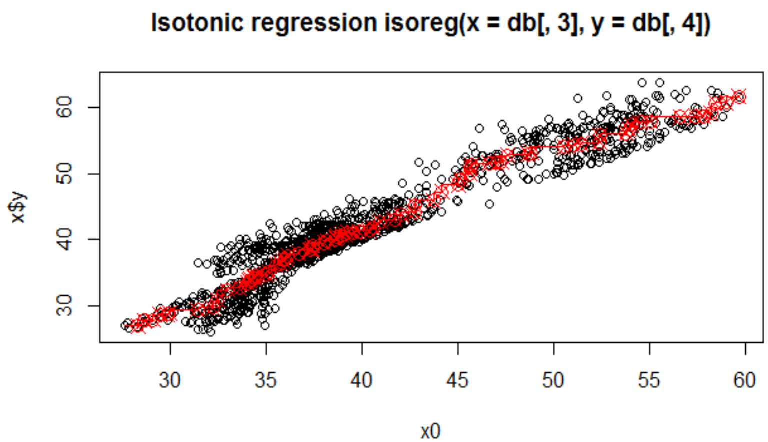
Minimize $\sum_i w_i(y_i - \hat{y}_i)^2$

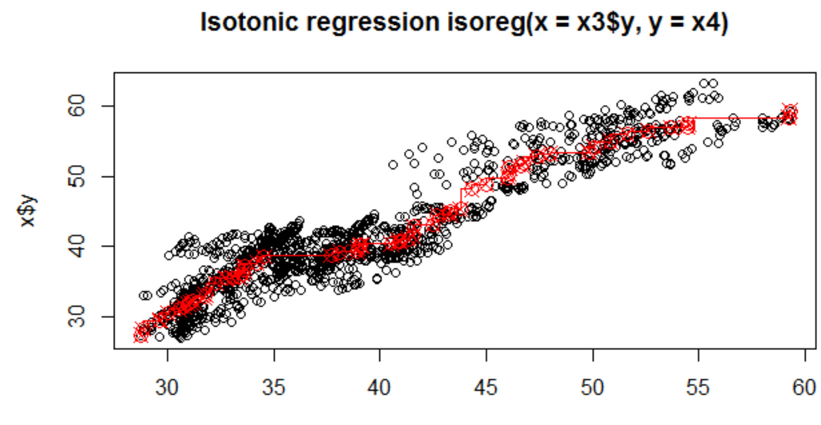
Subject to $\hat{y}_{min} = \hat{y}_1 \leq \hat{y}_2 \ldots \leq \hat{y}_n = \hat{y}_{max}$

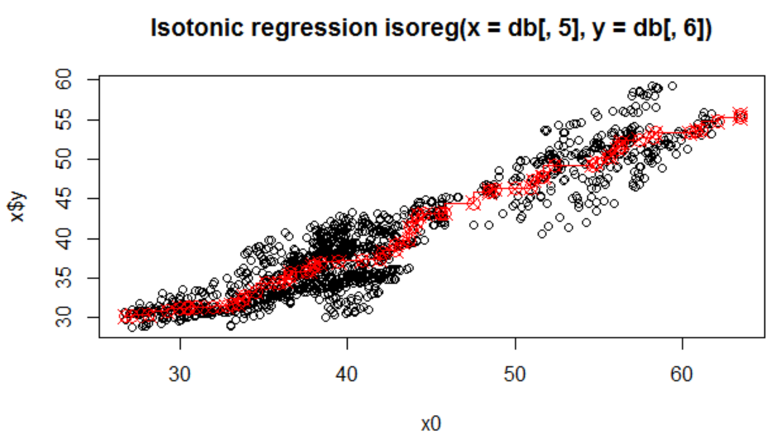
The isotonic regression finds a non-decreasing approximation of a function while minimizing the mean squared error on the training data.

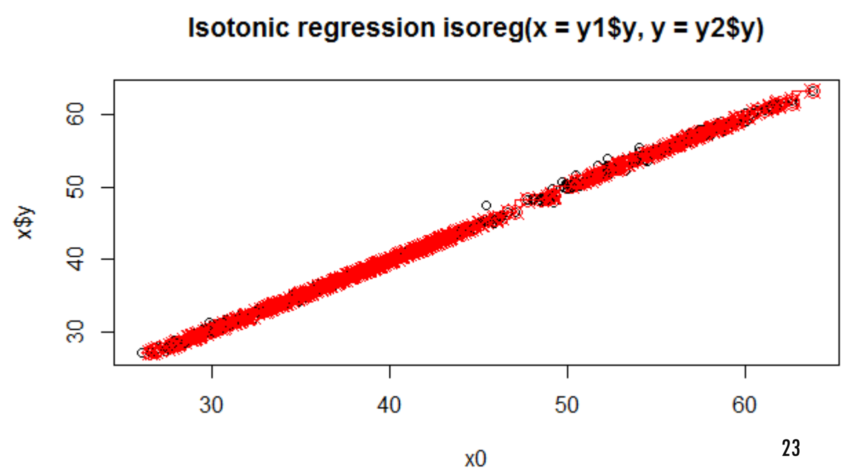

Isotonic regression isoreg(x = db[, 1], y = db[, 2])
x$y
x0


Isotonic regression isoreg(x = x1$y, y = x2$y)
x$y


Isotonic regression isoreg(x = db[, 3], y = db[, 4])
x$y
x0


Isotonic regression isoreg(x = x3$y, y = x4)
x$y


Isotonic regression isoreg(x = db[, 5], y = db[, 6])
x$y
x0


Isotonic regression isoreg(x = y1$y, y = y2$y)
x$y
x0

CHI-SQUARE VALUES

- Chi-squared = (observed-expected)^2/(expected)
- Lower the value of Chi square higher the chances of success.

Model No	Model Name	Chi-square Values
1	Polynomial Model	117.732204
2	Holt-Winters Model	8.199687
3	Feed Forward Model	203.591952
4	ARIMA Model	8.941534
5	SES Model	8.132946
6	Market-Stock Correlation	153.113115
7	Stock Autocorrelation	8.396684
8	Weighted average of predicted Models	8.051282
9	Isotonic Regression	8.466684

➢Disadvantages of RF

- Random forest is observed to overfit for datasets with noisy classification/regression tasks.
- Random forest is more helpful for the classification of the data and not for the prediction of the continuous data.
- For data including categorical variables with different number of levels, random forests are biased in favor of those attributes with more levels. Therefore, the variable importance scores from random forest are not reliable for this type of data.

NEWS SENTIMENT ANALYSIS TO PREDICT STOCK MARKET TRENDS

GET RICH!

➢ Acquiring The Text Streams:

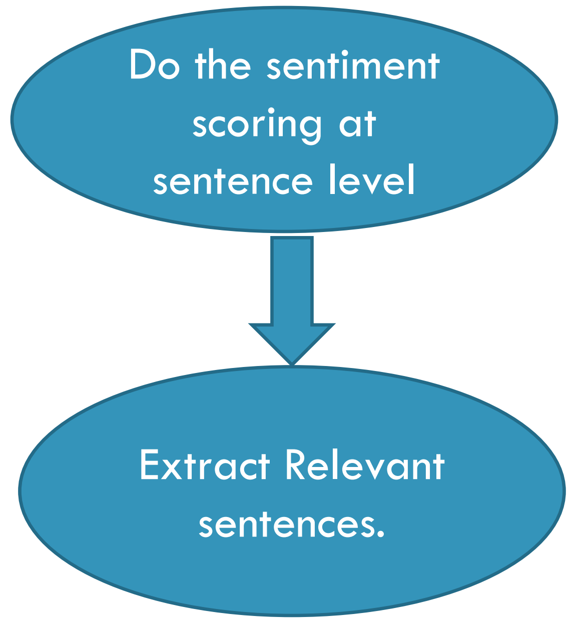
Few current streams of news items acquired from Google and Yahoo:

- GoogleFinanceSource ("NYSE:ACN")
- GoogleNewsSource (“ACN")
- YahooFinanceSource (“ACN")
- YahooInplaySource ()
- YahooNewsSource (“ACN")

Usage of webCorpus function for turning them into corpus objects.

3769 News stories captured today by the same (documents).

➢Breaking documents down into Sentences

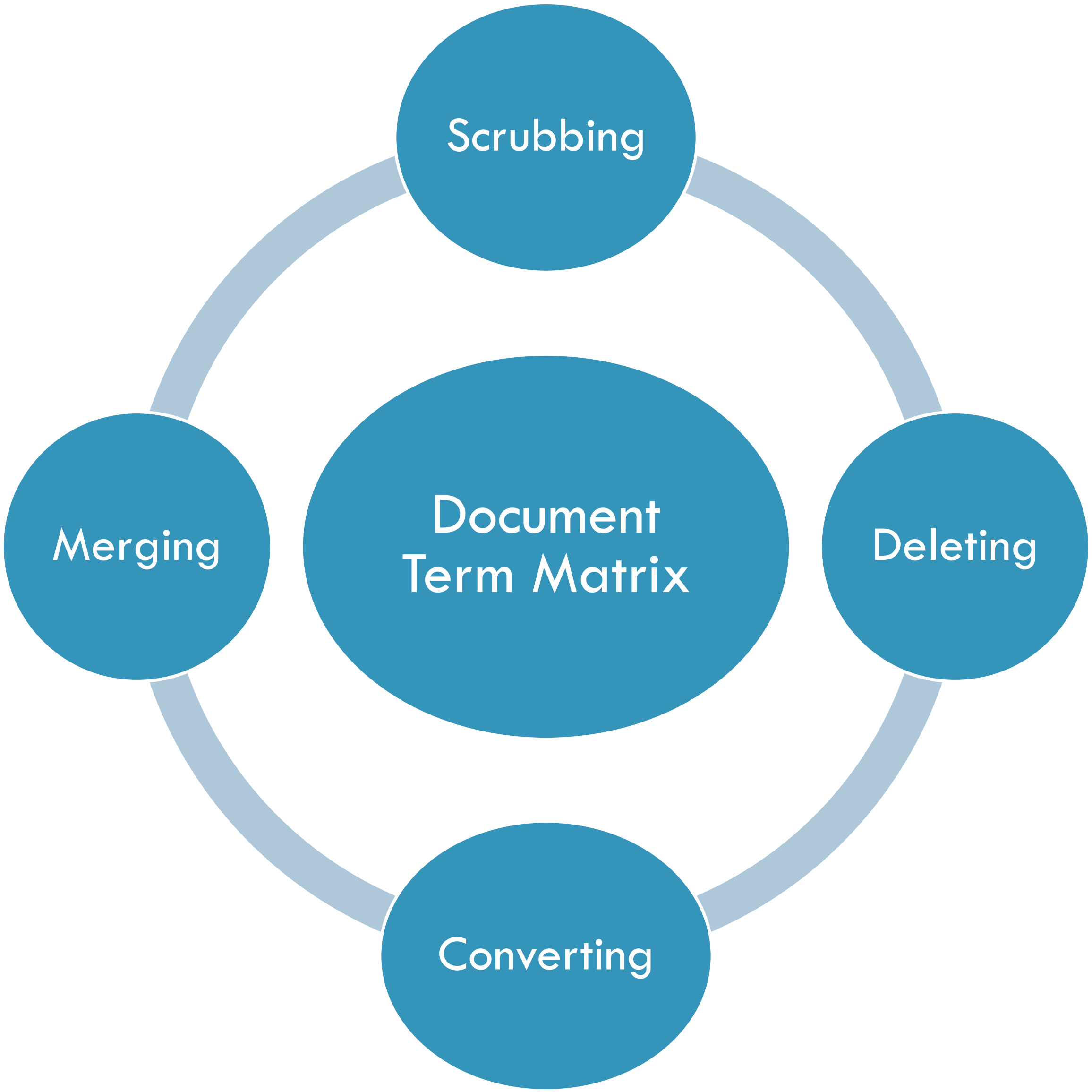


➢Focusing on relevant sentences instead of tagging the entire news story

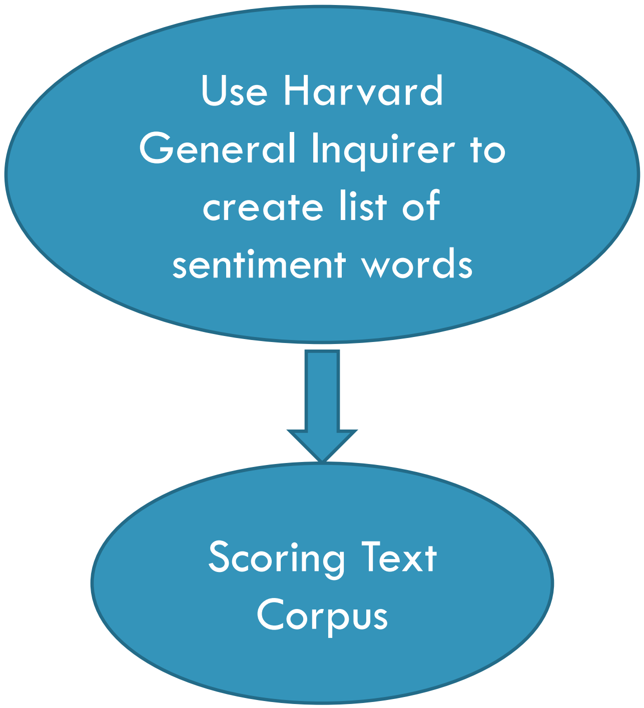
Apple Inc. (AAPL) sank 2.8 percent, the most since October, to $605.23. After rising to a record on April 9, the most valuable technology company fell for a fourth day in the longest losing streak since December.

Coinstar Inc. (CSTR) surged 7.3 percent to $65.78. The owner of the Redbox movie-rental kiosks said first-quarter sales and profit exceeded its previous projection and lifted its earnings forecast for 2012 to at least $4.40 a share.

➢Forming Document Term Matrix out of Corpora



➢Identifying Polarity of Words



For Example, the sentence:

- “ACN continues its phenomenal run” is a positive sentence as count(positive) = 2 and count(negative) = 0
- “Cracks develop in PCLN” is negative heading as count(positive) = 0 and count(negative) = 1

➢Visualization

POSITIVE WORD CLOUD

NEGATIVE WORD CLOUD

➢ Interpreting the Sentiment Score

Sentiment score should fall in the range 0-100%

- 23 negative terms occurring in the sentences
- 107 positive terms occurring in the sentences
- 89 sentences with sign +1
- 10 sentences with sign -1
- A naive sentiment score of 89 / (89 + 10) = 89 / 99 ≈ 89.89%

Sentiment Score = Positive instances / Total instances

Thank You