**Time Series Analysis on Stock Market Data –** **ARIMA model**

Predicting Stock Market Price Using ARIMA Model

**Abstract: 300 to 400 words**

### Introduction

The stock market has its biggest challenge of predicting the stock prices. The stock price data represents a financial time series data which becomes more difficult to predict due to its characteristics and dynamic nature. [1 Indian stock market prediction using artificial neural networks on tick data Dharmaraja Selvamuthu, Vineet Kumar & Abhishek Mishra]

### Case description

The use of Support Vector Machines (SVM) and Artificial Neural Networks (ANN) is wide in the prediction of stock market prices and its movements. The different methods of learning implemented by the algorithms gives a unique perspective for various insights. ARIMA Model method is a way to integrate technical analysis for making cognizant financial decision.

### Discussion and evaluation

Most ordinary techniques used in the prognosticate a financial time series are Support Vector Machine (SVM), Support Vector Regression (SVR) and Back Propagation Neural Network (BPNN). In the given abstract, we use the full historical data of a company, analyse it with algorithmic models like LSTM, ARIMA, ANN, Bayesian Regularization We use neural networks based on three different learning algorithms, i.e., Levenberg-Marquardt, Scaled Conjugate Gradient and Bayesian Regularization.

### Conclusion

The accuracy on the historical dataset comes out to be……..respectively, significantly less in accordance with the whole historical data.

**Introduction**

The introduction of your review should accomplish three things:

* **Introduce your topic: time series analysis**
  + It may sound redundant to "introduce" your topic in the introduction, but often times writer's fail to do so. Let the reader in on background information specific to the topic, define terms that may be unfamiliar to them, explain the scope of the discussion, and your purpose for writing the review.
* **State your topic's relevance: usage od ARIMA and other time series forecasting algorithms**
  + Think of your review paper as a statement in the larger conversation of your academic community. Your review is your way of entering into that conversation and it is important to briefly address why your review is relevant to the discussion. You may feel the relevance is obvious because you are so familiar with the topic, but your readers have not yet established that familiarity.
* **Reveal your thesis to the reader: plot of the best algorithm to be used**
  + The thesis is the main idea that you want to get across to your reader. your thesis should be a clear statement of what you intend to prove or illustrate by your review. By revealing your thesis in the introduction the reader knows what to expect in the rest of the paper.

Stock price prediction is one of the most widely studied and challenging problems, attracting researchers from many fields including economics, history, finance, mathematics, and computer science. The volatile nature of the stock market makes it difficult to apply simple time-series or regression techniques. Financial institutions and traders have created various proprietary models to try and beat the market for themselves or their clients, but rarely has anyone achieved consistently higher-than-average returns on investment. Nevertheless, the challenge of stock forecasting is so appealing because an improvement of just a few percentage points can increase profit by millions of dollars for these institutions. Traditionally, many prediction models have focused on linear statistical time series models such as ARIMA [7]. However, the variance underlying the movement of stocks and other assets makes linear techniques suboptimal, and non-linear models like ARCH tend to have lower predictive error [17]. Recently, researchers have turned to techniques in the computer science fields of big data and machine learning for stock price forecasting. These apply computational power to extend theories in mathematics and statistics. Machine learning algorithms use given data to “figure out” the solution to a given problem. Big data and machine learning techniques are also the basis for algorithmic and high-frequency trading routines used by financial institutions. In this paper we focus on a specific machine learning technique known as Support Vector Machines (SVM). Our goal is to use SVM at time t to predict whether a given stock’s price is higher or lower on day t +m. We look at the technology sector and 34 technology stocks in particular. We input four parameters to the model - the recent price volatility and momentum of the individual stock and of the technology sector. These parameters are calculated using daily closing prices for each stock from the years 2007 through 2014. We analyze whether this historical data can help us predict price direction. If the Efficient Markets Hypothesis (EMH) holds true, prices should follow a random walk and be unpredictable based on historical data. We find that in the short-term this holds true, but in the long-term we are able to reach prediction accuracies between 55% and 60%. We conclude that our model is able to achieve significant prediction accuracies with some parameters in the long-term, but that we must look at more granular intra-day trading data to achieve prediction accuracies in the short-term. The code written can be found at https://github.com/SaahilMadge/Spring-2015-IW.

**Discussion**

The discussion section is the body of your paper. The discussion section contains information that develops and supports your thesis. While there is no particular form that a discussion section must take there are several considerations that a writer must follow when building a discussion.

* **Don't summarize!**
  + A review paper is not simply a summary of literature you have reviewed. Be careful not to leave out your own analysis of the ideas presented in the literature. Synthesize the material from all the works—what are the connections you see, or the connections you are trying to illustrate, among your readings.
* **Analyze, Synthesize, Interpret.**
  + A review paper is not a pure summary of the information you read for your review. You are required to analyze, synthesize, and interpret the information you read in some meaningful way. It is not enough to simply present the material you have found, you must go beyond that and explain its relevance and significance to the topic at hand. Establish a clear thesis from the onset of your writing and examine which pieces of your reading help you in developing and supporting the ideas in your thesis.
* **Stay focused.**
  + Keep your discussion focused on your topic and more importantly your thesis. Don't let tangents or extraneous material get in the way of a concise, coherent discussion. A well focused paper is crucial in getting your message across to your reader.
* **Organize your points.**
  + Keeping your points organized makes it easier for the reader to follow along and make sense of your review. Start each paragraph with a topic sentence that relates back to your thesis. The headings used for this guide give you some idea of how to organize the overall paper, but as far as the discussion section goes use meaningful subheadings that relate to your content to organize your points.
* **Relate the discussion to your thesis.**
  + Your thesis should illustrate your objectives in writing the review and your discussion should serve to accomplish your objectives. Make sure your keep your discussion related to the thesis in order to meet your objectives. If you find that your discussion does not relate so much to your thesis, don't panic, you might want to revise your thesis instead of reworking the discussion.

**Conclusions**

The conclusion

Because the conclusions section often gets left for last it is often the weakest part of a student review paper. It is as crucial a part of the paper as any and should be treated as such.

A good conclusion should illustrate the key connections between your major points and your thesis as well as they key connections between your thesis and the broader discussion—what is the significance of your paper in a larger context? Make some *conclusions*—where have you arrived as a result of writing this paper?

Be careful not to present any new information in the conclusion section.

**References**

Here you report all the works you have cited in your paper. The format for a references page varies by discipline as does how you should cite your references within the paper.

**ARIMA:-**

basically means that the time series

needs to have a constant mean constant

variance over time has no seasonality

and so on

so this seems like it satisfies most of

those conditions except of course it

doesn't have a constant mean over time

the mean is of course shifting upward it

seems in a linear fashion so we can't

use this straight ARMA model which is a

shame because if we were to somehow

eliminate this trend then we probably

could use it because the rest of the

conditions would be satisfied that's

where the ARIMA model comes in so you

use the ARIMA model in situations like

this where things seem like they're

stationary except for a pesky moving

average moving mean

that comes in here so arima stands for

autoregressive moving average and the I

stands for integrated

in this context just means that instead

of predicting the time series itself

you're going to be predicting

differences of the times

from one timestamp to the previous

timestamp

the basic form of ARIMA model is

ARIMA 1 1 1

notice that arma had two parameters P

and Q the P being the AR order and the

to being the MA order ARIMA has three

parameters a P of D

the Q the P and Q are the same the P

applies to the worker of the AR part the

Q is the order of the MA part and the D

as you might have guessed is the order

of the integrated part in this case what

we did was a D equals 1 a difference 1

because we just took the first

difference

(formula)

the second difference which is you take

your Z of T series and you transform it

again which basically what that means as

you create a new series like W sub T

which is Z sub T plus 1 minus Z sub T so

you do a second difference we can do a

third difference usually it's the Feist

is just to do a first one but of course

it depends on the exact task you have at

hand so the simplest form of ARIMA is 1

1 1 and that's going to be given by the

mathematical model which is going to be

Z sub T that's the thing we're trying to

predict right no longer we're trying to

predict your anchors but rather the

difference between your anchor sails

from one time point to the last is equal

to V 1 Z sub T minus 1 this is the

autoregressive bit right because we're

doing Z sub T as a function of Z sub T

minus 1 plus theta 1 which is the

coefficient times epsilon t minus 1 this

is of course the moving average bit as

always and then we of course have our

error in the current time period so

looking at this equation the AR bit is

here the MA bit is here and the I bit

the integrated bit is taken care of by

the fact that Z sub T is a difference

between consecutive time points of the

series we originally started with ok

hopefully that makes sense now the last

question of course is let's say we form

a amazing model for Z sub T it's very

accurate we're happy with it of course

we don't really want in the end to

predict the difference but we want to

predict how many anchors are expected to

sell next month so the actual series so

how do we get back from Z sub T to a sub

T in order to make predictions in that

series well it's not too tough so how do

we recover a sub K let's suppose we have

a sub 0 a sub 1 all the way to a sub L

so let's say this last point time point

right here is 2 equals help and of

course the Y value here would be a sub

and let's say after that we don't have

any information that's what we want to

predict right so what we want is a sub k

which is at some point k in the future

and of course we want to figure out

what's the y-value which is gonna be a

sub k how many anchors did I sell at

that time point so we want a sub K now

if we just transform this equation right

here a sub K let's say T plus 1 is K you

shift this a sub T over to this side and

these T's are of course K minus once

because it's one less than the subscript

here so that means a sub K is equal to c

sub k plus z sub k minus 1 plus a sub k

minus 1 we just keep going a sub k minus

1 is equal to z sub K minus 2 plus a sub

K my say a sub K minus 2 and of course

we still had the Z sub K minus 1 here

and we keep going on and on and on and

on until what we eventually get is the

sum of all Z sub K minus I where I goes

from 1 to K minus L plus a sub L why did

we stop at E sub L because that's the

last a value that we actually had data

for so since we actually know that we

don't need to go any further okay so I

know that was a little bit mathy and if

you want please pause here and convince

yourself of this before going forward

but once you convince yourself of that

you see that this is going to give us

our prediction for a sub k because we

have a sub L that's the last recorded

value of anchor anchor sold and we have

all these Z sub K s because we use this

model to predict them so if we just do

all that addition we get our best guess

for a sub K number of anchors sold in

time period K okay

so just to recap the ARIMA model is not

that much crazier than just an ARMA

model it's really just used when the

time series you're trying to predict has

an obvious linear trend upward or

downward even for that for that matter

and then you would go ahead and take

first differences if you're using an

aroma with a 1 in the center for the

deep second difference is if you're

doing a 2 here and if you want to figure

out you know which differencing should I

use well it's basically when your time

series become stationary so if we did

first differences and we figure out the

plot still does not stationary we could

try a second difference and a third

difference but typically you want to

stick with as low of an order as

possible to keep your model simple okay

and then of course changing this P or Q

would be the same thing as an ARMA model

you would have more of these lags in the

AR you would have more of these air lags

and

May and that's what that would mean okay

and to recover your original time series

**LSTM:-**

This Recurrent Neural Network tutorial will help you understand what is a neural network, what are the popular neural networks, why we need recurrent neural network, what is a recurrent neural network, how does a RNN work, what is vanishing and exploding gradient problem, what is LSTM and you will also see a use case implementation of LSTM (Long short term memory). Neural networks used in Deep Learning consists of different layers connected to each other and work on the structure and functions of the human brain. It learns from huge volumes of data and used complex algorithms to train a neural net. The recurrent neural network works on the principle of saving the output of a layer and feeding this back to the input in order to predict the output of the layer. Now lets deep dive into this video and understand what is RNN and how does it actually work.

Transcript:-

short-term memory problem

LSTM is a special version of RNN

which solves the short-term memory

problem and in this video I will explain

LSTM

in a very simple manner using real-life

examples

let's say you have NLP task for centers

completion

this particular network layer which

looks like this. so there are set of

neurons

in that layer and this hidden state is

nothing but a short-term memory.

this hidden state is actually

containing

the short-term memory. now if you want to

remember

long-term memory we need to introduce

another state called

long-term memory. so that state is called

C so there are two states now hidden

state which is short-term memory

and the there is a self-state which is a

long-term memory.

in traditional RNN if

it looks something like this so I have

drawn the vertical neurons here but

you can draw on draw them this

in a horizontal fashion as well so it's

just a layer of neurons and your x(t) and

h(t) are vectors

so when you have a word for example you

will first convert into a vector vector

is nothing but a

list of numbers and your hidden state

will be

also a vector and using both these

vectors you will do

you know like sigma operation like

weighted multiplication and then you

apply

activation function which is which is

tan h in the case of RNN

and then you get a new hidden state so

**SVM:-**

Support Vector Machines are powerful supervised learning algorithms for both classification and regression. It is a discriminative classifier that is formally defined by a separating hyperplane. So given labelled training data, the algorithm outputs an optimal hyperplane that categorizes new examples.

Or   
Transcript :-

support vector machine a lot

of times referred to as the SVM

.

applications of the support vector

machine at least some general ones that

are commonly used with it face detection

text in hypertext categorization

classification of images and

bioinformatics these are only about a

few of those that are used with this SVM

.

six different sections we're gonna start

with what is machine learning so we can

see where the vector machine fits in why

the support vector machine what is a

support vector machine and understanding

.

support vector fits in under

classification deciding what yes-and-no

is and there is also a regression

version but it is primarily used for

classification let's take a detour and

see if we can connect us to the human

experience and find out why support

vector machine

the support vector machine once we go

through an understanding of how it works

and what it looks like we're gonna look

at the advantages of support vector

machine and finally dive into a use case

in Python will write some script on it

.

support vector machine model SVM is a

supervised learning method that looks at

data and sorts it into one of two

categories

.

advantage of the support

vector machine

we'll start with high dimensional input

space or sometimes referred to as the

curse of dimensionality we looked at

earlier one dimension to dimension three

dimension when you get to a thousand

dimensions a lot of problems start

occurring with most algorithms that have

to be adjusted for the SVM automatically

does it in high dimensional space one of

the high dimensional space one high

dimensional space that we work on is

sparse document vectors this is where we

tokenize the words in document so we can

run our machine learning algorithms over

though I've seen ones get as high as 2.4

million different tokens that's a lot of

vectors to look at and finally we have

regularization parameter the realization

parameter or lambda is a parameter that

helps figure out whether we're going to

have a bias or overfitting of the data

whether it's going to be over fitted to

a specific instance or is gonna be

biased to a higher low value with the

SVM it naturally avoids the overfitting

and bias problems that we see in many

other algorithms these three advantages

of the support vector machine make it a

very powerful tool to add to your

repertoire of machine learning tools