Attrasoft Predictor for Windows
Version 2015.0

User's Guide and Reference Manual
http://attrasoft.com

Attrasoft
P. O. Box 13051
Savannah, GA. 31406
gina@attrasoft.com

Statement of Copyright Restriction

The Attrasoft program that you have purchased is copyrighted by Attrasoft, and your rights of ownership are subject to the limitations and restrictions imposed by the copyright laws outlined below.

It is against the law to copy, reproduce or transmit (including, without limitation, electronic transmission over any network) any part of the program except as permitted by the copyright act of the United States (title 17, United States code). However, you are permitted by law to write the contents of the program into the Machine memory of your computer so that the program may be executed. You are also permitted by law to make a back-up copy of the program subject to the following restrictions:

- Each back-up copy must be treated in the same way as the original copy purchased from Attrasoft;
- No copy (original, or back-up) may be used while another copy, (original, or back-up) is in use;
- If you ever sell or give away the original copy of the program, all back-up copies must also be sold or given to the same person, or destroyed;

This User's Guide and Reference Manual is copyrighted by Attrasoft.

Order Information

Attrasoft Predictor for Windows

Online order:
      http://attrasoft.com
      gina@attrasoft.com

Mail order:
      Predictor
      Attrasoft, Attn.: Gina
      P. O. Box 13051
      Savannah, GA. 31406
      USA

http://attrasoft.com
gina@attrasoft.com

About Attrasoft Predictor

Making decisions based on your projection of the future events is a way of life. **Attrasoft Predictor** is a tool for your data. It uses a sequence of numbers to predict the next row of numbers in line. It uses historical data to forecast future possibilities. It does not matter what you want to predict.

Attrasoft Predictor can be used for:

- Predict Stock Market
- Predict earnings and revenue of a company
- Predict short term and long term interest rates
- Predict commodity (gold, oil, corn,...) prices
- Predict regional/worldwide price fluctuations for a particular merchandise item
- Predict various indicators/indexes of the economy for strategic thinking and policy issues
- Predict Dynamic Systems
- Predict Markov Chains
- Predict other complex systems

…

Customized versions of the **Predictor** are also available.

Installing Attrasoft Predictor for Windows

To setup:
>	Create a folder
>	Copy CD:*.* to the folder.

To Run from C: drive:
>	Click the only exe file in the folder.

© Attrasoft 2015

Table of Contents

1. Introduction

1.1 Why Predictor

Making decisions based on your projection of the future events is a way of life. How do you make this kind of projection? Most likely you will make a prediction based on your historical experience.

In general, a decision is made based on several factors. Some of them can be measured and have a historical track record.

This is where software can help you to predict these factors based on your historical data. The software can learn years of experience from your data and use the knowledge to tell you what will happen if this were in the past. This provides you with a baseline for your decision making process. The software can learn years of experiences in seconds and is ready to serve you as if it is one of your staff.

1.2 Why Attrasoft Predictor
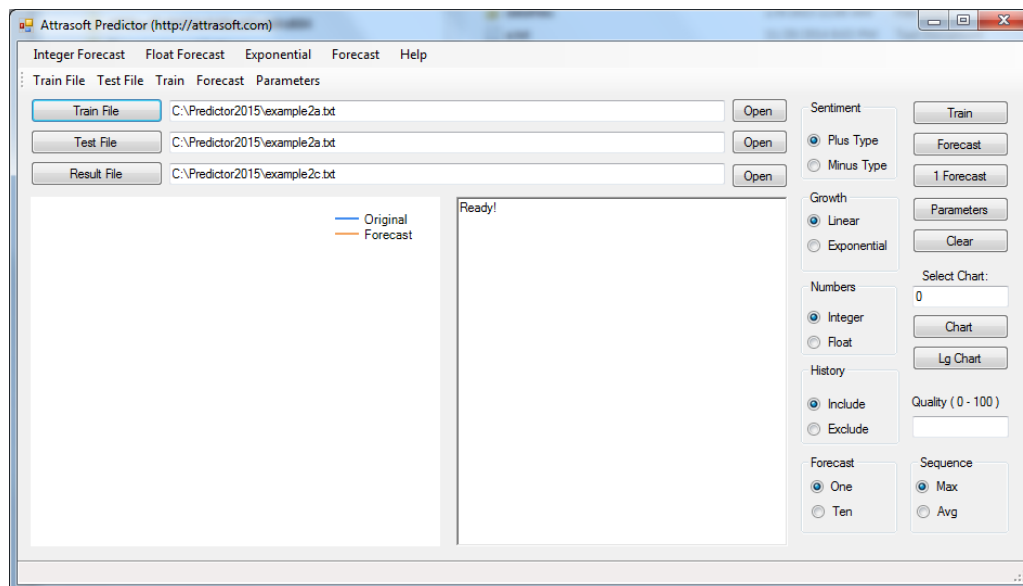
Attrasoft Predictor is a fast data processing tool for your data. It uses a sequence of numbers to predict the next row of numbers in line. It uses historical data to predict future possibilities. It does not matter what you want to predict.

99% of the work to use the **Predictor**, is to prepare your data containing your historical data. The remaining 1% of the work is to operate the **Predictor** (four clicks). The **Predictor** is especially good if you have a large amount of data because of its accuracy and speed.

The advantages of the **Attrasoft Predictor** are:

- The **Predictor** is <u>universal</u>: it does not matter what you want to predict.
- The **Predictor** is <u>scalable</u>: it is designed to handle large databases.

- The **Predictor** is <u>fast</u>: it is designed to handle huge volumes of data in a way that you can make decisions fast.
- The **Predictor** is very <u>simple</u> to operate: your data is 99% of the work required to use the **Predictor**. If you already have the data, then 99% of the work is done. 4 clicks are all that is required for the **Predictor** to present you with a forecast.

Once your data is prepared correctly, the **Predictor** is able to provide you with rated predictions on any subject or any problem. **Attrasoft Predictor** can:

- Predict Stock Market
- Predict earnings and revenue of a company
- Predict short term and long term interest rates
- Predict commodity (gold, oil, corn,...) prices
- Predict regional/worldwide price fluctuations for a particular merchandise item
- Predict various indicators/indexes of the economy for strategic thinking and policy issues
- Predict Dynamic Systems
- Predict Markov Chains
- Predict other complex systems

…

1.3   How to Operate Attrasoft Predictor

The **Attrasoft Predictor** makes patterns out of complicated problems. The **Predictor** is based on the neural network technology developed at Attrasoft. A neural net learns from past experience; you have to train the software first. You have to teach the **Predictor** by showing it your historical data.

 The operation has two phases:

1.   Your raw data
   - Set up a data set (a set of variables that describes your problem)
   - Collect data
   - Preprocessing if necessary

2.   Using the Predictor
   - Prepare the Train file and Test file
   - Link the Train file to the Predictor
   - Link the Test file to the Predictor
   - Train
   - Forecast

1.4   Parameters

The software is a black box. The black box has a few parameters to setup. The most important parameters are:

- Data set;

- Historical data;
- Trend;
- Precision level.

If you are the beginner, you can ignore the following parameters. The additional parameters are:

- Sentiment (Bullish or bearish)
- Growth Type (Linear or Exponential)
- Number Type (Integer or float)
- Forecast History (Include or Exclude)
- Number Of Forecasts (Forecast one row or many rows)
- Sequence Type (Use most likely outcome or average outcome)
- Blurring (How much error the underlying neural net can have, 0 - 100)
- Sensitivity (How sensitive the underlying neural net can have to small variations, 0 - 100)
- Threshold (When the matching score is below the threshold, it will be ignored)

There are also a set of measurements that will help you understand your prediction:

- Prediction Quality Score
- Prediction Score
- Prediction Distribution (possibilities and associated probabilities)

There are a huge amount of parameters not exposed to users. They can only be set through ordering customized Predictor software.

1.5   Predictor's Version

The **Predictor** has several versions:

Version	Neurons	prices
Standard version	10,000	$499.99
100K version	100,000	$999
250K version	250,000	$1,499
1M version	1,000,000	$2,999

2.   Getting Started

We will start with an existing example. There are two data files:

- Training file
- Test File

The Training file contains the entire history of data. The Test file contains the last 100 rows of data, so the Predictor can judge the performance.

To use the Predictor:

1. Start the Predictor

2. Click the Open (Train File) button.

This click will open the Train file already linked to the Predictor.
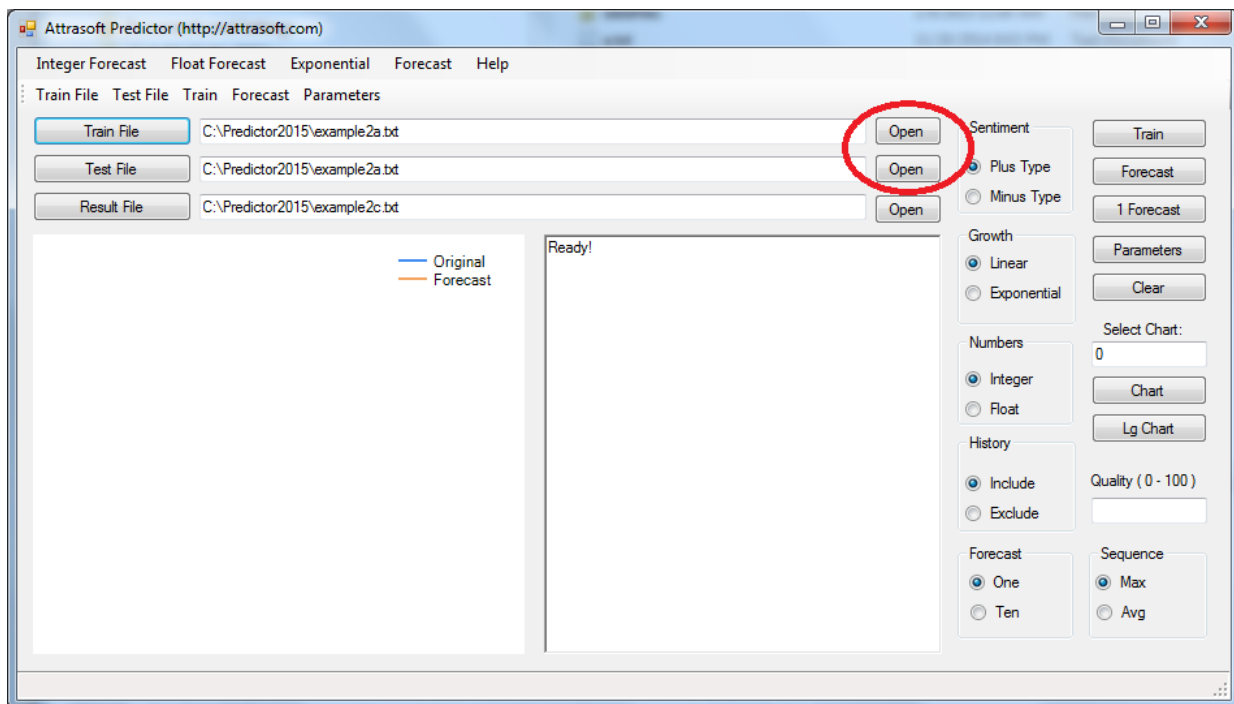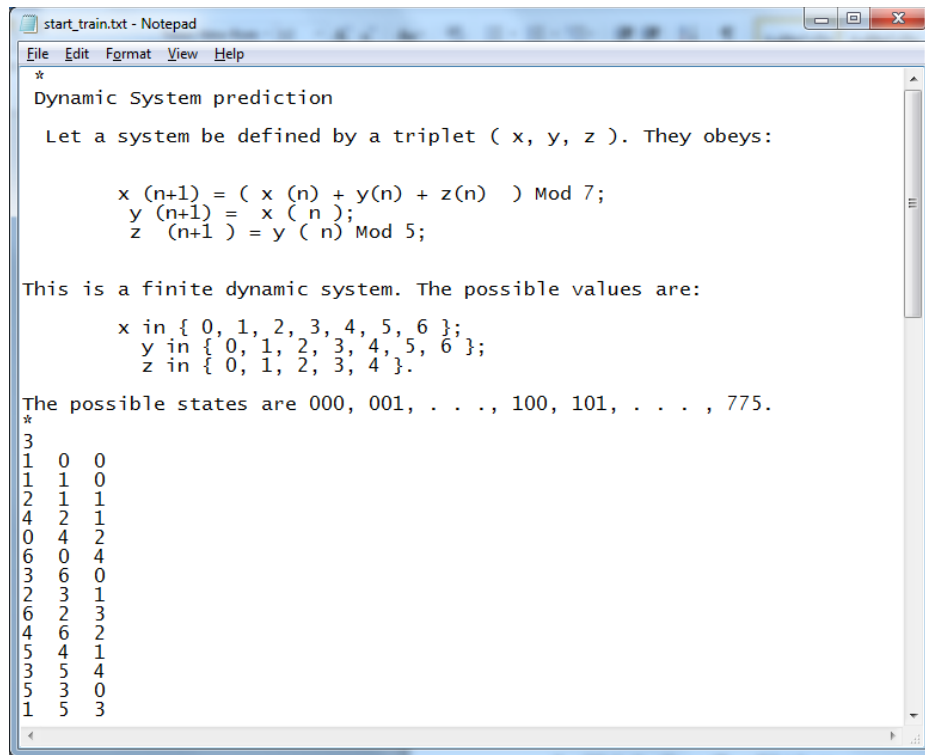

The file has a remarks section:

*
…
*

This section will be ignored by software. This file has three columns of data:

1 0 0

start_train.txt - Notepad

File   Edit   Format   View   Help

```
*
Dynamic System prediction

 Let a system be defined by a triplet ( x, y, z ). They obeys:


        x (n+1) = ( x (n) + y(n) + z(n)  ) Mod 7;
        y (n+1) =  x ( n );
        z   (n+1 ) = y ( n) Mod 5;


This is a finite dynamic system. The possible values are:

        x in { 0, 1, 2, 3, 4, 5, 6 };
        y in { 0, 1, 2, 3, 4, 5, 6 };
        z in { 0, 1, 2, 3, 4 }.

The possible states are 000, 001, . . ., 100, 101, . . . , 775.
*
3
1   0  0
1   1  0
2   1  1
4   2  1
0   4  2
6   0  4
3   6  0
2   3  1
6   2  3
4   6  2
5   4  1
3   5  4
5   3  0
1   5  3
```
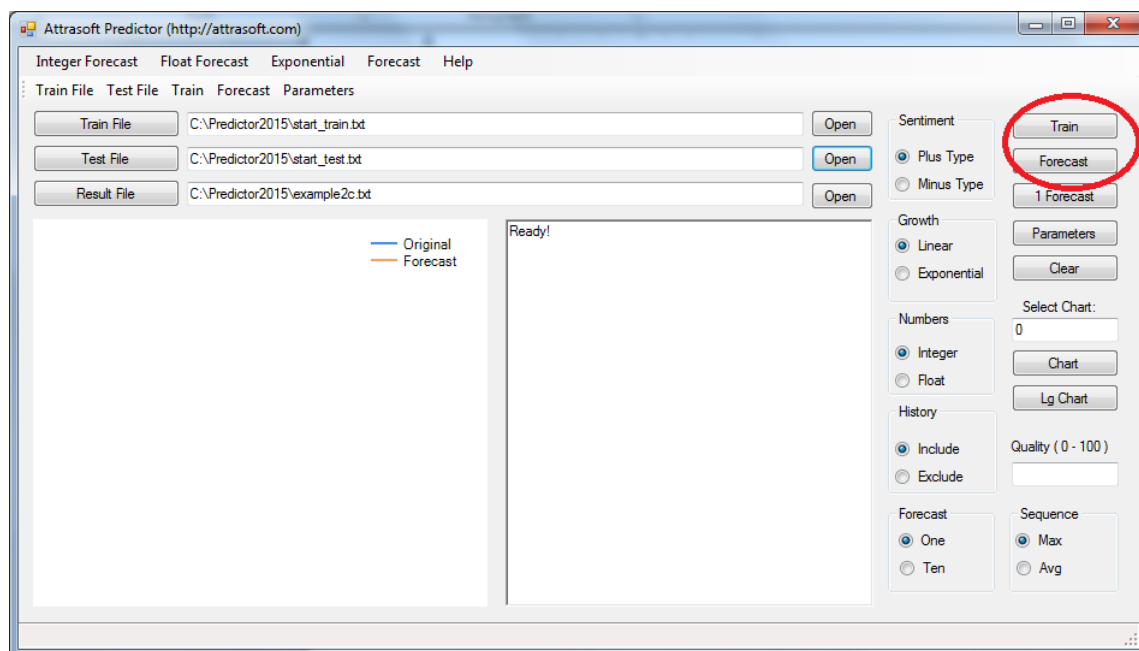
3. Click the Open (Test File) button.

This click will open the Test file already linked to the Predictor.

```
start_test.txt - Notepad
File   Edit   Format   View   Help

 *
 Dynamic System prediction

   Let a system be defined by a triplet ( x, y, z ). They obeys:


         x (n+1) = ( x (n) + y(n) + z(n)  ) Mod 7;
          y (n+1) =   x ( n );
          z   (n+1 ) = y ( n) Mod 5;

This is a finite dynamic system. The possible values are:

         x in { 0, 1, 2, 3, 4, 5, 6 };
          y in { 0, 1, 2, 3, 4, 5, 6 };
          z in { 0, 1, 2, 3, 4 }.

The possible states are 000, 001, . . ., 100, 101, . . . , 775.
 *
3
1   0   0
1   1   0
2   1   1
4   2   1
0   4   2
6   0   4
3   6   0
2   3   1
6   2   3
4   6   2
5   4   1
3   5   4
5   3   0
1   5   3
```

4. Click the Train button

5. Click the Forecast button

After the computation ends, you will see the forecasted next row:

4 6 4.

6. Evaluation

The above graph indicates that the Quality score of the prediction is 100, meaning, it is the best the Predictor can do.

In the following, we can use the chart to compare forecasted results vs the real data for the three columns (0, 1, 2). The 0 in the text box indicates that column 0 is currently selected. Click the Lg Chart button (Large Chart) to open the large chart below. The figure below indicates that the 'predicted results' and the 'real results' are the same, which is why the quality score is 100.
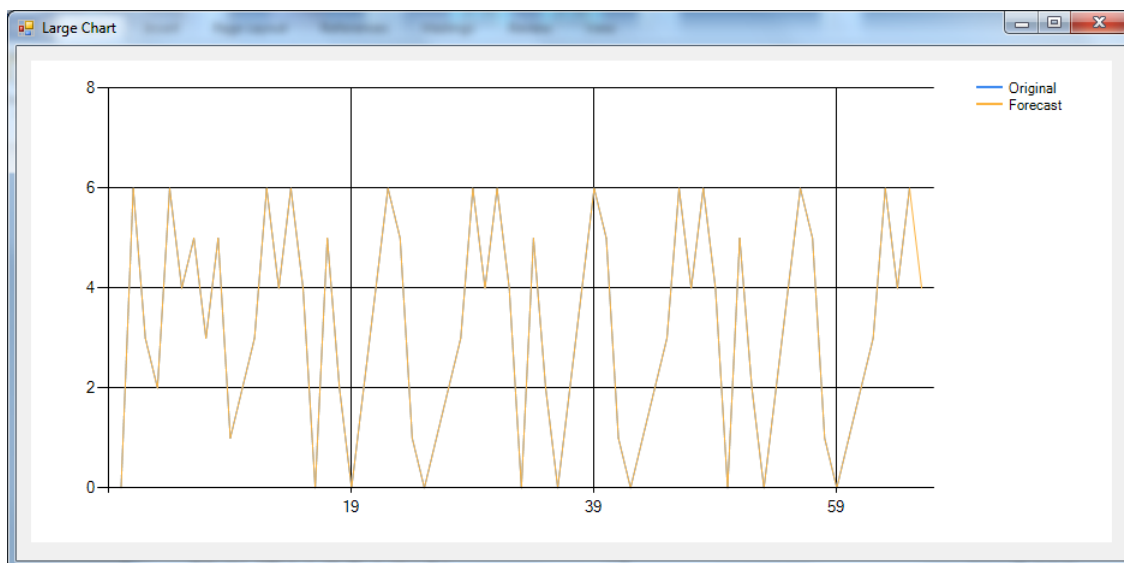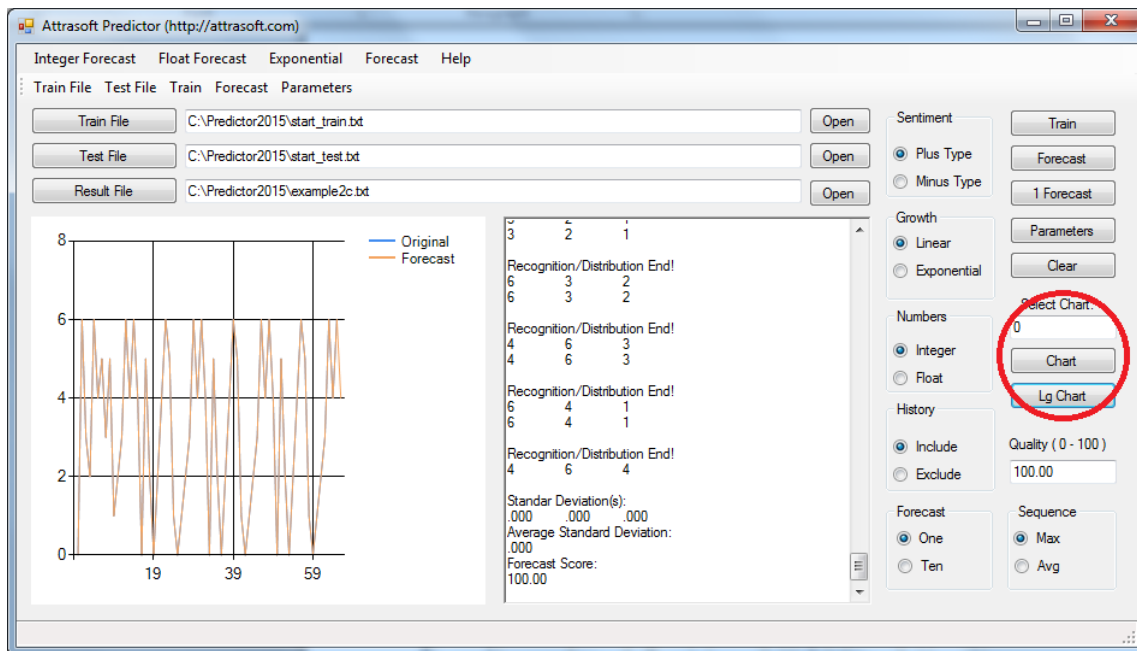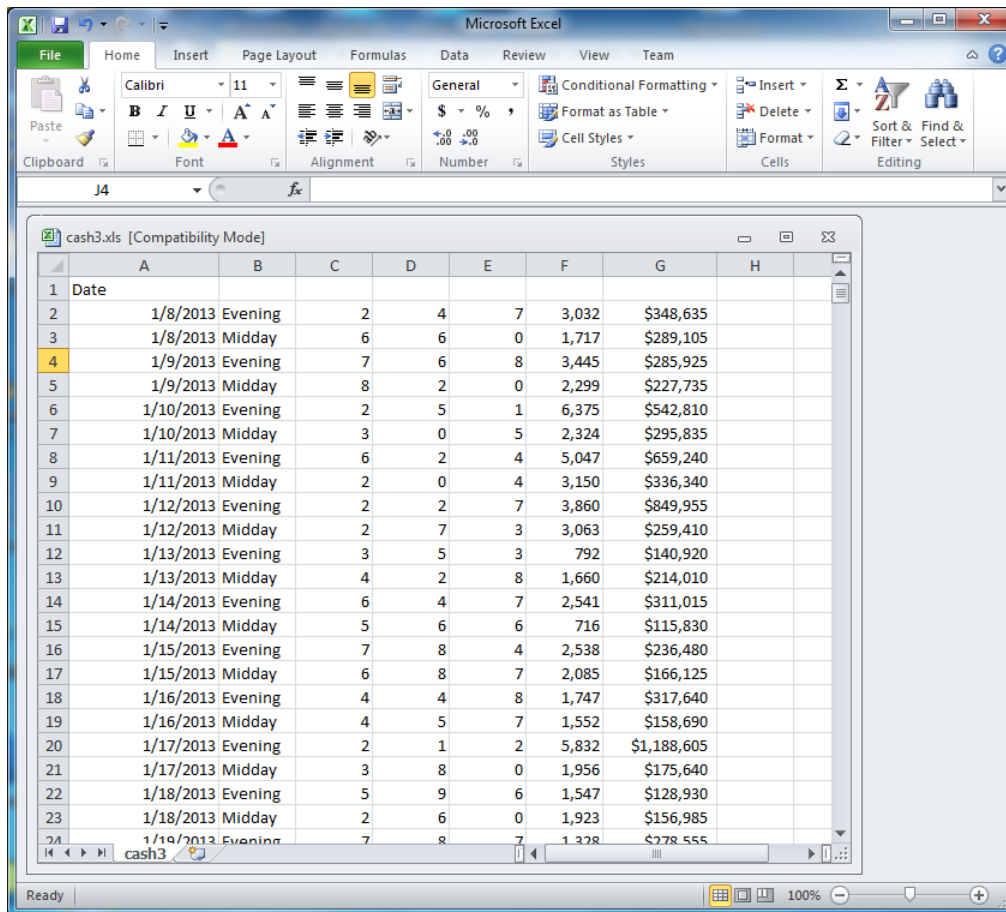
3.   How To

People usually use the Predictor over and over again. We recommend that you save your data in Excel. To make a prediction.

1. **Organize your data in a spreadsheet.**

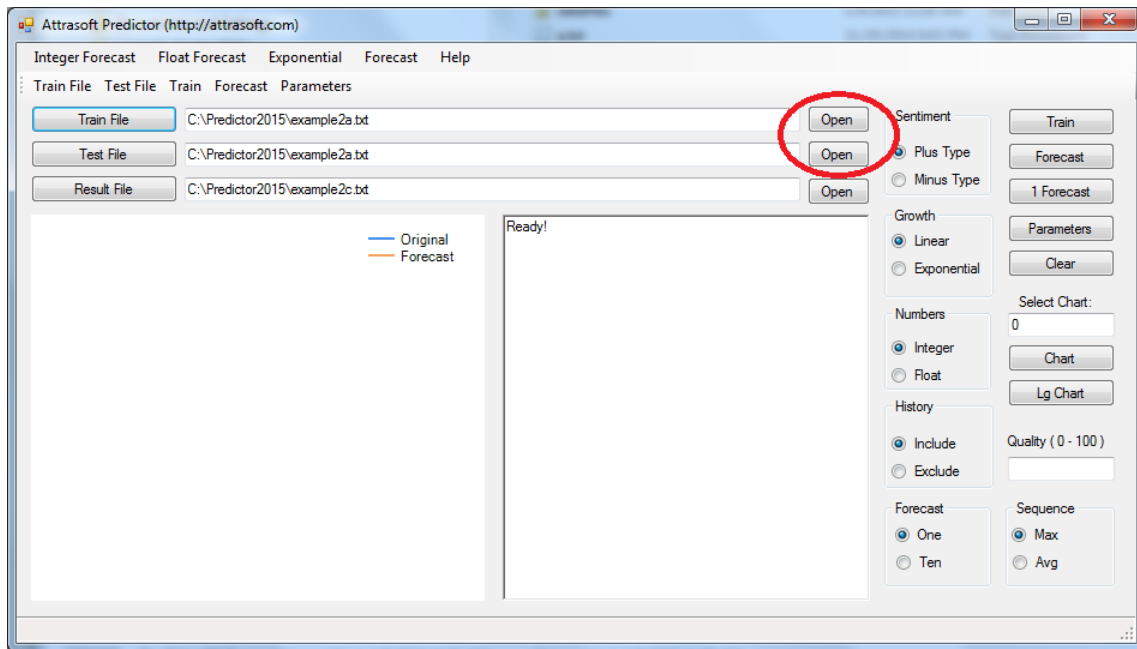   Assume the following spread sheet is our historical data:

2. Start Predictor.

3. Click the Open (Train File) button.

This click will open the Text file already linked to the Predictor.

Now click "Edit/Select All" and Delete key to get rid of old data.

4. Enter Data.

4.1 Enter the number of columns.

Assuming our problem has 1 column, enter:

*
*
1

4.2 Copy & Paste the data.

Copy 1 column from the spread sheet:

Paste the data to the input file:

```
*
*
1
2
6
7
8
2
3
6
2
2
2
3
4
6
5
7
6
4
4
2
3
5
5
2
7
1
8
9
0
6
7
1
4
```
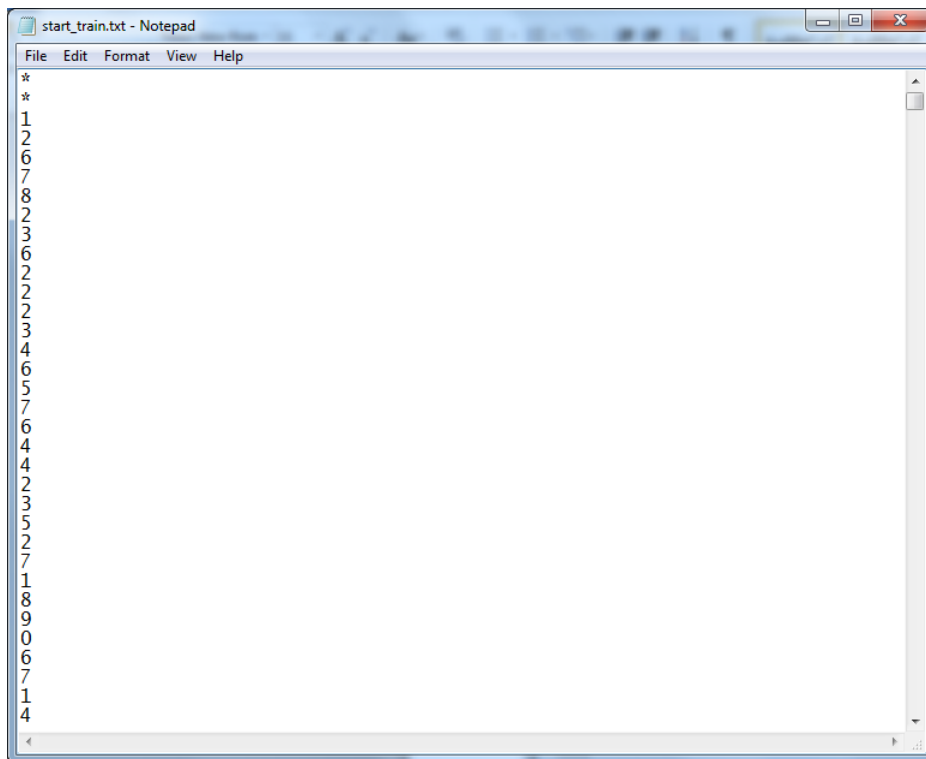
Click the "Save" button to save the data.

5. Click the Open (Test File) button.

This click will open the Test file already linked to the Predictor. Enter:

*
*
1

Now, we cut and paste the last 100 or so columns to this file:

start_test.txt - Notepad

File Edit Format View Help

```
*
*
1
7
2
0
4
9
6
8
0
2
5
1
6
9
5
5
4
1
6
9
0
0
4
9
5
3
3
0
7
4
6
7
```
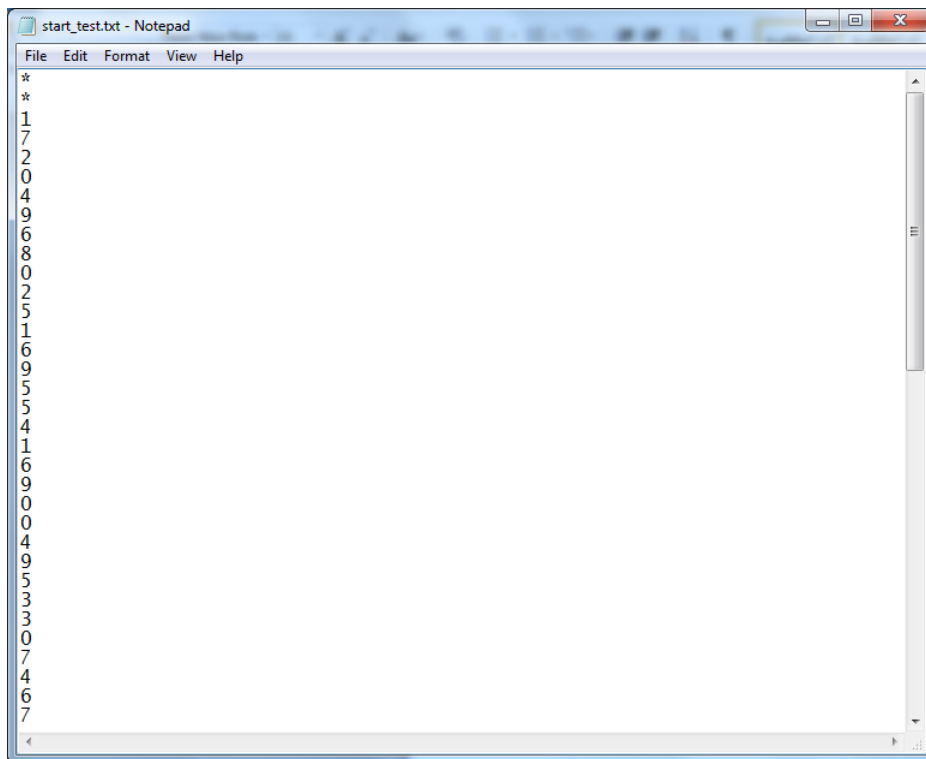
6. Click the Train Button

7. Click the Forecast button

You will get the results:
0
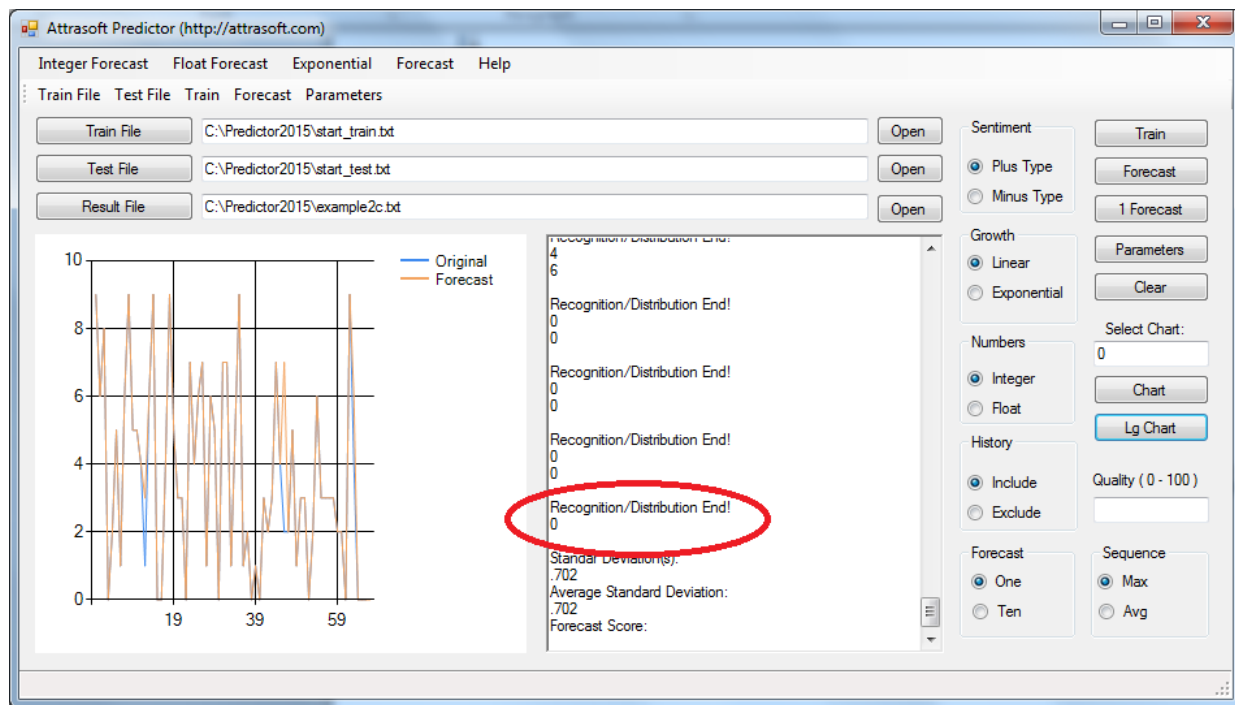
Click the Lg Chart button, you will get a chart comparison:
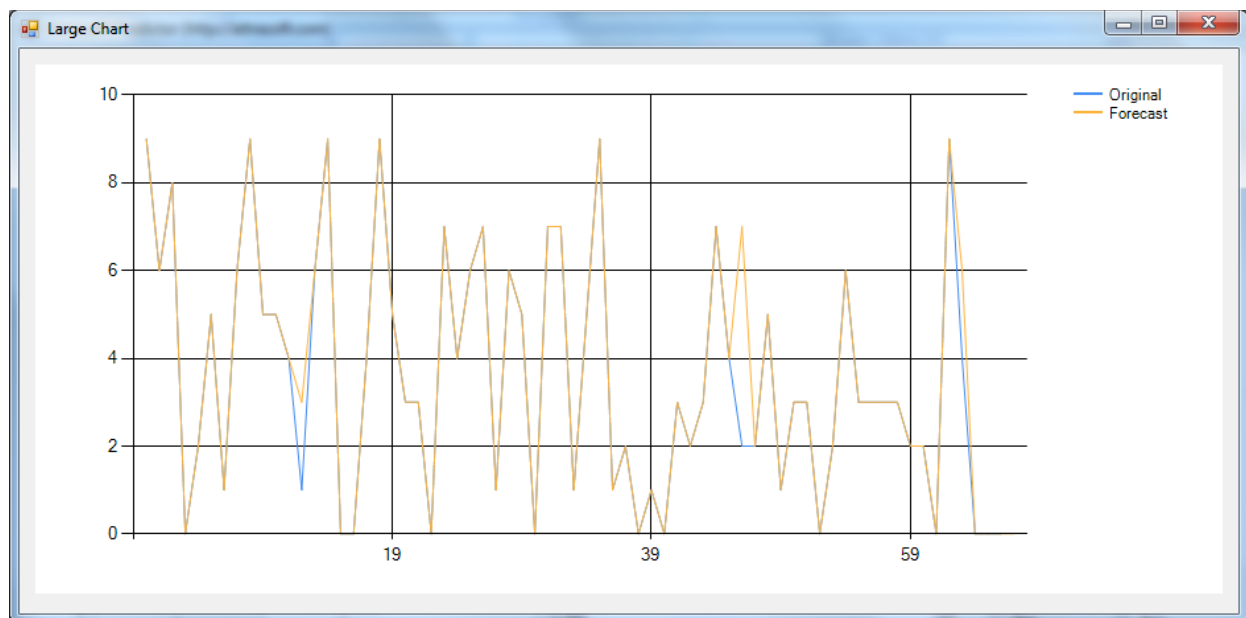
4. Parameters

There are several parameters in the user interface:
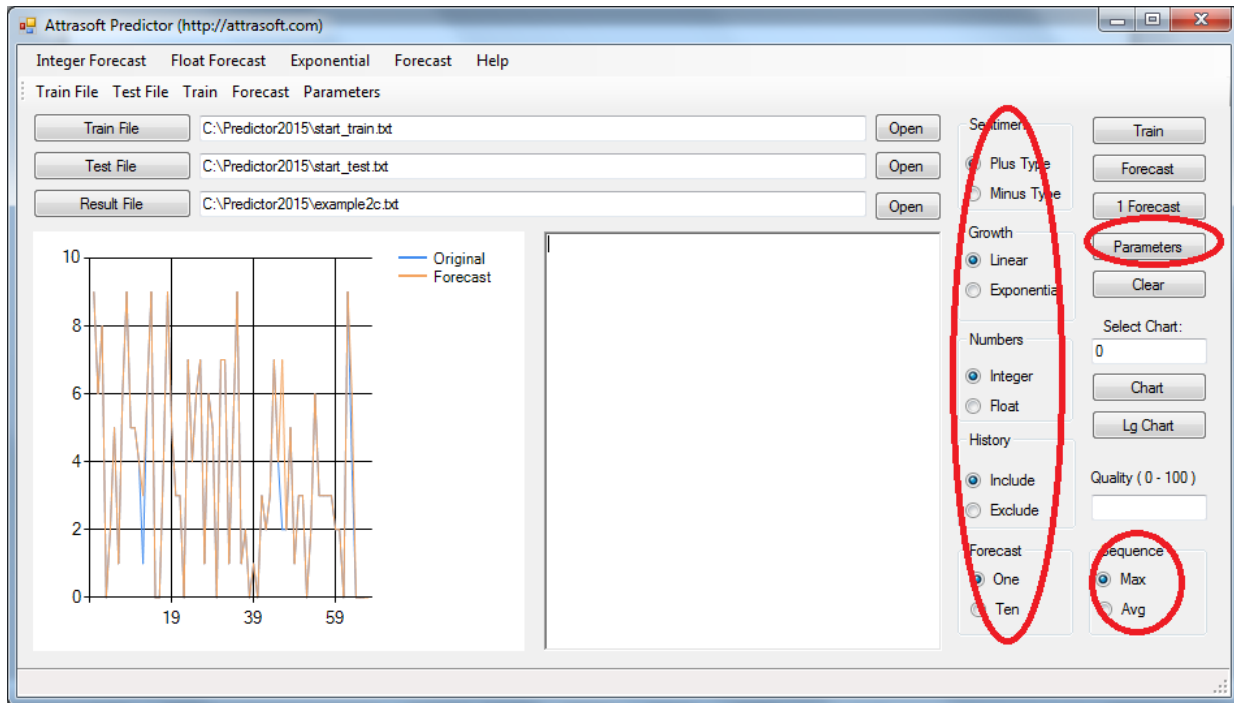
The Parameter button will open more parameters in the figure below.
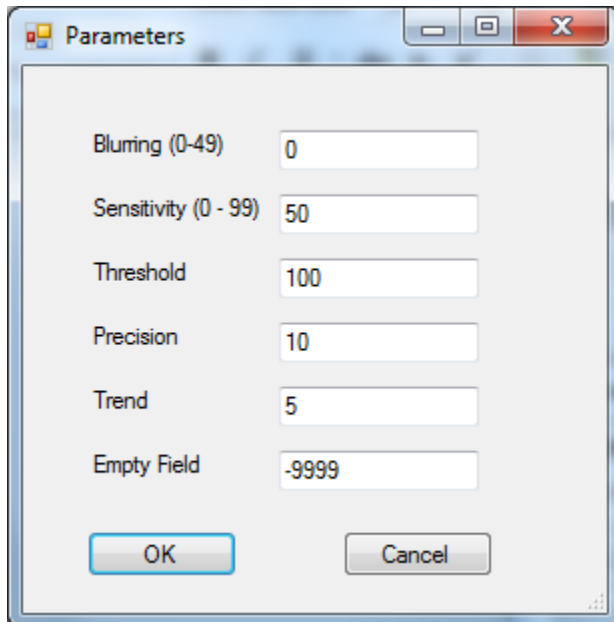
Each parameter will change the forecast results. In addition, the data set and the historical data will also change the predictions.

4.1 Data

The Historical data can change the forecast. There are two variations:

- Data set
- Volume

From your historical database, you will determine a data set, which consists of the relevant variables for the prediction. Then get the data for these variables; if necessary, preprocess them, such as calculating the moving-average.

The **Predictor** predicts the values of the next period or next several periods. Take a stock data as an example,

- If you have daily data, you will predict next day's data;
- If you have weekly data, you will predict next week's data;
- If you have monthly data, you will predict next month's data;
- ...

The amount of data also changes the forecast results. For example, 1 year data may generate a different result from 10 years data.

4.2 Growth Type

This parameter has two values:

- Linear
- Exponential

The Linear mode deals with two situations:

- Data is in fixed intervals [a, b];
- Data grows linearly, 1, 2, 3, 4, … .

The Exponential mode deals with:

- Data grows exponentially, 1, 2, 4, 8, … .

4.3 Sentiment Type

This parameter has two values:

- plus
- minus

Here, "plus" and "minus" means the higher forecast and the lower forecast, respectively.

4.4   Trend

What is an N-trend? Suppose you have 5 rows of data:

3
4
7
2
1,

then a 2-trend is the combination of all 2 consecutive numbers together:

(3 4), (4 7), (7 2), (2 1).

A 3-trend is:

(3 4 7), (4 7 2), (7 2 1).

A 4-trend is:

(3 4 7 2), (4 7 2 1).

N-trend means the **Predictor** looks at N rows of data at a time. This is an important parameter used by the **Predictor**. The N-trend is a user-selected variable. If you do not make a choice, the default value is 5.

To change the Trend:

- click "Parameter";
- enter your number.

Under similar prediction confidence numbers, the longer the Trend, the more accurate the prediction will be. You have to balance the two sides of this parameter. On one hand, a 10-trend certainly provides a more accurate prediction than a 5-trend. On the other hand, a 10-trend prediction requires a much larger volume of data. One way to increase N is to use preprocessed data. Using the moving-averages, can increase N significantly.

4.5   Precision Level

The Precision-level determines the error of each variable. The higher the Precision is, the lower the errors are. This is one of the important parameters used by the **Predictor**. The Precision level is a user-selected variable. If you do not make a choice, the default value is 10.

This parameter will directly change the errors in the output file. To change the Precision:

- click "Parameter";
- enter your number.

You have to balance the two sides of this parameter. On one hand, a 100-level prediction certainly provides more precision than a 10-level prediction. On the other hand, a 100-level prediction requires a much larger volume of data. One way to increase the precision level is to use preprocessed data. Using the moving average, can increase the precision level significantly.

In general, the longer the trend, the more accurate the prediction will be. The higher the precision, the lower the error(s) will be.

Between these two factors, the **Trend** is more important than **Precision**. If possible, always goes for the maximum Trend.

Here is how to set the **Trend**: as you increase the **Trend** from low to high, initially, you can see the confidence (rating) increase. If you keep increasing the **Trend**, for most problems, at certain point the confidence begins to drop. You should establish an acceptable confidence level and that will tell you when to stop.

4.6 Roles of the Four Parameters

The four parameters have the following impact on the prediction:

- Data **volume**: one year's data will produce more accurate results than one month's data;
- Number of variables in the **data set**: more variables requires more data;
- **Trend**: a longer trend requires more data than a shorter trend;
- **Precision level**: a higher precision prediction requires more data than a lower one.

If you run the **Predictor**, and get too much information, you can do one or more of the following:

- Reduce the data volume;
- Increase the number of variables in the data set;
- Increase the Trend;
- Increase the Precision level.

If you run the **Predictor**, and get no results, you can do one or more of the following:

- Data preprocessing;
- Increase the data volume;
- Reduce the number of variables in the data set;
- Reduce the Trend;
- Reduce the Precision level.

Data preprocessing can improve the result significantly. There are noise fluctuations in the data. These noises can be removed from the data by a simple procedure like introducing the moving averages.

However, if the noise fluctuation is left in the data, the **Predictor** will think these noise fluctuations are meant to be learned and will treat the noise fluctuation as well-defined patterns in the data. As a result, it will shorten the Trend and reduce the Precision.

5.   Command Reference

5.1   Files

"Train File"
>   Use the "Train File" command to select the train file.

"Train File" Open
>   Use the "Train File Open" command to open the train file.

"Test File"
>   Use the "Test File" command to select the test file.

"Test File" Open
>   Use the "Test File Open" command to open the test file.

"Result File"
>   Use the "Result File" command to select the result file.

"Result File" Open
>   Use the result File Open" command to open the result file.

5.2   GUI Parameters

The GUI variables are:

- Sentiment (Bullish or bearish)
- Growth Type (Linear or Exponential)
- Number Type (Integer or float)
- Including History (Include or Exclude)
- Forecast Numbers (Forecast one row or many rows)
- Sequence Type (Use most likely(max) outcome or average outcome)

Growth Type

This parameter has two values:

- Linear
- Exponential

The Linear mode deals with two situations:

- Data is in fixed intervals [a, b];
- Data grows linearly, 1, 2, 3, 4, … .

The Exponential mode deals with:

- Data grows exponentially, 1, 2, 4, 8, … .

Sentiment Type

This parameter has two values:

- plus
- minus

Here, "plus" and "minus" means the higher forecast and the lower forecast, respectively.

Number Type

This parameter has two values:

- Integer
- Real

Some sequences are integer sequences; integer sequences use integer forecast.

5.3   Integer/Real/Exponential

Integer/Train
Train the software.

Integer/1 Forecast (Max, History)
The history of real data and forecast data (the forecast with maximum score) are compared and the forecast scores are given at the end. The forecast results are the last forecast rows.

Integer/1 Forecast (Avg, History)
The history of real data and forecast data (the average forecast) are compared and the forecast scores are given. The forecast results are the last forecast rows.

Integer/1 Forecast (No History)
The forecast results are presented. The results list all possibilities and their scores.

Integer/10 Forecast (Max, History)
The history of real data and forecast data (the forecast with maximum score) are compared and the forecast scores are given at the end. The forecast results are the last 10 forecast rows.

Integer/1 Forecast (Avg, History)
The history of real data and forecast data (the average forecast) are compared and the forecast scores are given at the end. The forecast results are the last 10 forecast rows.

Integer/10 Forecast (Max, No History)
The 10 max forecast results are presented.

Integer/10 Forecast (Avg, No History)
    The 10 average forecast results are presented.

Appendix A. Data Set

A.1   Data Set

Let us start from the beginning:

A "**data set**" is a set of relevant factors related to your prediction problem.

We will use the stock market as an example.

What are the factors for the stock market? i.e. what is a **data set** for a stock? There are hundreds of them:

- Interest Rates
- Short term: 3-month US Treasury Bills
- Intermediate term: 10-year US Treasury Bills
- Long term: 30-year US Treasury Bills
- ...

- Inflation
- Inflation rate
- Consumer Price Index
- Producer Price Index
- ...

- Economy
- GDP (Gross Domestic Product) Growth Rates
- Leading Economic Indicators, a key measure of future economy activity
- National Association of Purchasing Management Index
- Imports/Exports
- ...

- Stock Market
- Dow Jones Industrial Average (large)
- NASDAQ Composite (technology)
- S&P 500
- Russell 2000 (small)
- New highs/new lows
- Volumes
- ...

- Stock-Specific
- Stock value
- PE-ratio (Earnings should not be used because of the discontinuity: it changes once every 3 months)

- 200-day Moving Averages
- 100-day Moving Averages
- Stock/(100-day Moving Averages) ratio
- Section Indicator (for Example, phil.semecond.index)
- PR ratio (Price/Revenue, revenue should not be used because of the discontinuity: it changes once every 3 months)
- ...

- Social
- Sentiment index
- Population
- Social media topics
- Social media ranking
- ...

An example of a **data set** for **Intel** stock is:

- 3-month US Treasury Bills
- 30-year US Treasury Bills
- Consumer Price Index
- GDP (Gross Domestic Product) Growth Rates
- Dow Jones Industrial Average
- NASDAQ Composite
- Intel Stock value
- PE-ratio

A.2   Data Set Selection

Data preparation for a prediction is the most important factor in the prediction or forecast. Data preparation means two things:

- data set; and
- the number of rows.

Your data preparation directly influences the results. In this section, we will show you how to prepare data.

The better you choose the variables, the better the potential for the prediction. This is the single most productive area to work on. You can increase the information content of your data in several ways:

Adding variables: the first approach is to add variables to the data set. Assuming that a new variable captures some aspect of the problem not already included in the other variables you've chosen to be in the data set, the information content is increased. On one hand, obviously, you should continue to do so until you think all aspects of the problem are covered. Otherwise, the data set will not be complete. On the other hand, each additional variable increases the complexity of the problem exponentially. To compensate for this, more data is required. If your data is limited, then your ability to make a prediction is limited.

Data Preprocessing: This approach transforms the existing data into another form. Typically, this is just a mathematical manipulation of the data. A moving average, for example, is a mathematical manipulation of the underlying price variable. This approach may allow us to extract information more effectively or more efficiently. **For stock prediction, we always recommend you to use the moving averages of the underlying stock price.**

Removing Unnecessary Variables: This approach removes data that does not increase the information content. Overloading the **Predictor** with useless or redundant information will increase the complexity of the computation exponentially and will require more data. If you do not have enough "rows" for the additional variables, the quality of the prediction can be reduced. Averaging Intel and Microsoft stock prices, for example, is a mathematical manipulation, which covers both factors with only one number.

Increase Data Volume: This approach increases the data volume. Two years of data, for example, is better than one year of data, if the conditions in these two years are basically the same. In reality, there is a limit to the data volume. For the **Attrasoft Predictor**, always go for the maximum amount of available data. The cost of increasing data volume in computation time is almost 0. The data volume is not a factor in the **Attrasoft Predictor**, for the software is designed for terabyte processing.

Using similar data: It is quite possible that sometimes you have no data. In this case, you might consider using similar data. We will present an example in which, despite not having Intel data, we still want to predict Intel stock by substituting Intel with Microsoft, because Microsoft stock is similar to Intel stock. As you will see, the **Predictor** will still be able to make a direct hit.

Once again, data set selection is your job. This is the single most productive area to work in, regardless of the technology employed for your data-mining effort. Your largest contribution to your prediction is right here.

A.3   Data Preprocessing Using Excel

Data preprocessing transforms the existing data into another form. Typically, this is just a mathematical manipulation of the data. A moving average, for example, is a mathematical manipulation of the underlying price variable. This approach may allow us to extract information more effectively or more efficiently.

Appendix B.   Customized Software

Customized software can be ordered from Attrasoft upon your request for the following reasons:

- **Larger External Networks**: The current version supports up to 10,000 external neurons. If you want a larger one, it has to be customized. We produce neural networks up to 1,000,000 neurons.
- **Larger Internal Networks**: The current version supports up to 10,000 internal neurons. If you want a larger one, it has to be customized.
- **Special Data Interface**: If you need to interface the **Predictor** with your application data, a customized software can be ordered. Example, missing data, string data, ...
- **Software library**: you can develop your own predictors.

To get updated information:

http://attrasoft.com