PROJECT

*Analiza și proiectarea << ... >>*

|  |  |
| --- | --- |
| **Soos Andreea** | **C:\Users\ANDREEA\AppData\Local\Microsoft\Windows\INetCache\Content.Word\11042121_393034007524518_1450177677_n.jpg** |

# 

# Part I – Analysis Report

Introduction

In this thesis we will put forward all the steps that must be completed to obtain E-FAST, a financial application which will help us predict the trend of certain stocks based on the way they behaved in the past. Given the fact that no investment is a sure thing, trying to programmatically predict the trend of a certain stock using distributed process and working with historical data and moving averages might be the 21st century innovation in assisting financial decisions.

The importance of an application like this is huge, given the fact that the number of market participants in the world of stock investment is growing and that the financial market is becoming excessively unstable, defined by **irrational behavior** and crashes.

Even if there are lots of opportunities when it comes to investing, choosing one of them is a struggle. Every financial analyst and broker is trying to attract more and more people by assuring them that they have found the best allocation for a share and that they are using the up to top technique to find what that best allocation is.

E-FAST will be one of the up to top techniques because it will execute, faster than any other application, two of the most important indicators: the moving averages (Exponential Moving Average & Simple Moving Average).

Moving averages are used to gauge the direction of the current trend. Every type of moving average is a mathematical result that is calculated by averaging a number of past data points. Once determined, the resulting average is then plotted onto a chart in order to allow traders to look at smoothed data rather than focusing on the day-to-day price fluctuations that are inherent in all financial markets.

A simple moving average that is calculated by adding the closing price of the security for a number of time periods and then dividing this total by the number of time periods. Short-term averages respond quickly to changes in the price of the underlying, while long-term averages are slow to react.

Exponential Moving Average is a type of moving average that is similar to a simple moving average, except that more weight is given to the latest data. The exponential moving average is also known as "exponentially weighted moving average".

Along with SMA & EMA, evolutionary algorithm will be used. By bringing these indicators together, the results will be more accurate because they will be based on more complex data.

1. Identification and description of the problem

Not a lot of people spend enough time monitoring their investments, but they are rather “sleepy” when it comes to their stock portfolio. That’s why they miss a lot of things such as the abnormal levels of investment performance volatility during various periods of the market cycle.

But let’s explain first what volatility is. “Volatility refers to the amount of uncertainty or risk about the size of changes in a security's value. A higher volatility means that a security's value can potentially be spread out over a larger range of values. This means that the price of the security can change dramatically over a short time period in either direction. A lower volatility means that a security's value does not fluctuate dramatically, but changes in value at a steady pace over a period of time.” [1]

“Volatility can be measured in many ways but we will explain the much accurate and easier way to do it. Through a process known as the historical method, risk can be captured and analyzed in a more informative manner than through the use of standard deviation. To utilize this method, investors simply need to graph the historical performance of their investments, by generating a chart known as a histogram.” [2]

“Even though using the historical method has its valuable pros, most investors choose to use standard deviation. Standard deviation is simply defined as the square root of the average squared deviation of the data from its mean. While this statistic is relatively easy to calculate, the assumptions behind its interpretation are more complex, which in turn raises concern about its accuracy. As a result, there is a certain level of skepticism surrounding its validity as an accurate measure of risk.” [3]

1.1 Motivation

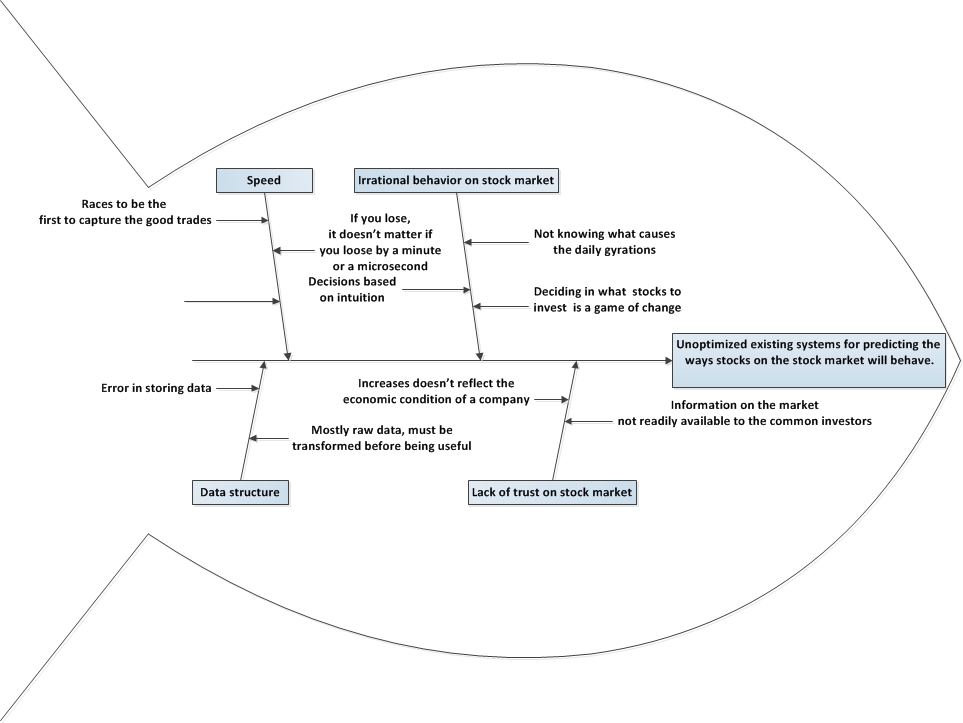


Fig 2.1

The saying ‘no investment on the stock market is a sure thing’ is probably the most common saying regarding the stock market. Our application’s purpose is to help the investor find the best strategy for a traded title. Using much simpler words, E-FAST will help the financial analyst find the allocation of a title with the maximum or close to maximum reward and with the minimum of risk.

Stock market is a “complicated” place where things move really fast, where data fluctuates and change its structure daily and where human’s error might alter the accuracy of data. Here, things have a tendency to go irrational and even if daily gyrations are common things that occur, their direction is hard to predict. Given all this facts, people started to lose their trust on the financial markets and as a result they invest less. Having assistance and guidance while deciding whether to invest or not, professional guidance, might be that encouragement people need to decide to invest more. This is what our financial application’s purpose is: offering support and assistance in daily decisions whether to trade or not.

Using moving averages and genetic algorithm, E-FAST is different from other applications in its field by offering improved accuracy and performance. Its main preoccupations are orientated towards making all the processes more efficient. Compared to other applications, what E-FAST brings new is the use of distributed system and concurrent algorithm along with the Moving Averages. This means speed, accuracy, faster access to data, so faster decisions.

2. Context

The stock market is trading over sixty trillion euros each year and this amount of money is greater than the price of goods and se­rvices traded on the entire world economy, so there is no need to explain why trying to bring improved applications in this domain it’s a must.

We will document our system’s context by using the facets approach.

* 1. The subject facet

Our main stakeholder is the financial analyst. A financial analyst is a person who performs financial analysis for clients, as part of their job. In our situation, he will be the one studying with vigilance the output of our application. Stakeholders can also be investment advisors, brokerage firms or any other person who has expertise in evaluating investments.

This application will work with data from different sources:

* Our first source of data will be our clients, who will provide a file containing information such as: stock names, values, data on which the stocks had this values and other necessary information for the result he (client) want to generate. The data from this file will be used by our application to obtain smoothed data, data that is easier to interpret by the financial analyst. Our application is not focused on day-to-day fluctuations from the financial markets but it works with this type of data too and if data fluctuated or not, will reflect in the interpretation of the results.
* Information might also come from different online financial data vendors.

Simple Moving Average is the most basic of the moving averages used for trading. The simple moving average formula is calculated by taking the average closing price of a stock over the last “x” periods. Trading based only on simple moving average is not encouraged by investors, but it happens sometimes if the stocks are braking out up or down strongly. Even if SMA was used long before EMA and it is easier to calculate and understand, it doesn’t place different importance on data and this one of their bigger limitation.

Exponential Moving Average is widely used, built on SMA framework and it places greater importance on more recent data: the larger the time period, the lower the importance of the most recent data. In order to calculate EMA, we need to get historic stock prices and calculate SMA first. Then, we will need additional data, such as: practiced commission for the type of share, the Exponential factor value, short and long terms for the MA number of days and information about the title : name, starting day of the analyzed period and the closing day of the analyzed period. To calculate an EMA, we have the following formula:

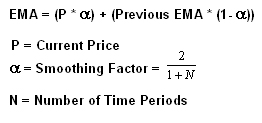


Fig 2.2

2.2 The functional facet

Our main stakeholder will be the financial analyst, who will use our application in order to obtain, for one or more of his shares, 2 of the most used and important indicators on the financial markets, moving averages: EMA & SMA.

For the financial analyst, using our application will be as easy as falling off a log. He will only have to upload a file with data and then choose carefully the number of days on which he wants the results. After that, all that is left to do is run the application.

After the execution, the analyst can choose whether to have the output on a text file or to continually use our application to obtain charts or diagrams with the data from the output.

* 1. The IT facet

What E-FAST brings new is the concept of grid computing. Comparing to distributed systems, where within the same network computers share one or more resources, in grid computing, computers share every resource and the computer network become one powerful supercomputer, where every computer have access to enormous processing power and storage capacity.

A grid computing system uses that same concept: share the load across multiple computers to complete tasks more efficiently and quickly. Computer’s resources are:

* Central processing unit (CPU): A CPU is a microprocessor that performs mathematical operations and directs data to different memory locations. Computers can have more than one CPU.
* Memory: In general, a computer's memory is a kind of temporary electronic storage. Memory keeps relevant data close at hand for the microprocessor. Without memory, the microprocessor would have to search and retrieve data from a more permanent storage device such as a hard disk drive.
* Storage: In grid computing terms, storage refers to permanent data storage devices like hard disk drives or databases.

Grid computing systems link computer resources together in a way that lets someone use one computer to access and leverage the collected power of all the computers in the system. To the individual user, it's as if the user's computer has transformed into a supercomputer.

By using grid computing, our application will finish calculations that cost lots of time faster, which means that, comparing to other application, given the same amount of time, we will be able to perform more calculations and have more executions.

We will use technologies such as JAVA & JDBC on NetBeans platform. Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation.[14] Java applications are typically compiled to byte code that can run on any Java virtual machine (JVM) regardless of computer architecture. JDBC is a Java database connectivity technology (Java Standard Edition platform) from Oracle Corporation. This technology is an API for the Java programming language that defines how a client may access a database. It provides methods for querying and updating data in a database. JDBC is oriented towards relational databases. A JDBC-to-ODBC bridge enables connections to any ODBC-accessible data source in the JVM host environment.

Our database will be in POSTGRESQL. PostgreSQL, often simply "Postgres", is an object-relational database management system (ORDBMS) with an emphasis on extensibility and standards-compliance. As a database server, its primary function is to store data securely, supporting best practices, and to allow for retrieval at the request of other software applications. It can handle workloads ranging from small single-machine applications to large Internet-facing applications with many concurrent users. Recent versions also provide replication of the database itself for availability and scalability.

* 1. Development facet

For developing this project, we worked according to Agile methodology. Agile development is a different way of managing IT development teams and projects. Comparing to Waterfall model, where every element results from the previous element, until all the phases of the projects are completed (Conception, Initiation, Analysis, Design, Construction, Testing and Deployment), Agile is like taking all the phases, mixing them in a blender and every new phase will be a mix of all the other.

Agile imply having a good communication between the stakeholder and the team and also stand-up meetings in which the tasks (user stories) for the next period of time are established. This period of time is called Sprint and it can last between one and four weeks, depending on the project. We will have as many sprints as needed until we finish the implementation of the project.

Having a good communication with the stakeholder comes with the benefit of having all the implementations regarding the Moving Averages and the Genetic Algorithm tested gradually and approved by the financial analyst himself.

Firstly, we divided our release schedule into a series of development iteration for two to four weeks. Planning involved scheduling the work to be done during an iteration and assigning individual work items to members of the team. The first iteration was creating the database. Second, we had to write code that takes a file with no extension received from a server, add the ‘zip’ extension and then extract content from that zip file. Third iteration was parsing that content from the zip file into the database. Forth iteration was building an interface for our application.

1. System requirements

What E-FAST aims to achieve is faster execution in calculating EMA & SMA, two of the most important indicators on the financial markets. Given the fact that the markets moves really fast and prices always fluctuate, it might happen that by the time you decide whether to invest or not, the information on witch you should’ve based your decisions has already been changed. With E-FAST you will get the information you need in the shortest amount of time.

E-FAST is also recommended by the fact that the calculations will have almost no errors for the numbers decimals points. This means that you can’t get results more accurate than what our E-FAST will give you.