Investment and Financial Forecasting
- A Data Mining Approach on Port Industry

Serkan Güneş

School of Computing
Blekinge Institute of Technology
Soft Center
SE – 372 25 Ronneby
SWEDEN
This thesis is submitted to the Department of Systems and Software Engineering, School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:
Author(s):
Serkan Güneş
E-mail: serkangunes@gmail.com

University advisor(s):
Dr. Lawrence Edward Henesey
School of Computing

School of Computing
Blekinge Institute of Technology
Soft Center
SE – 372 25 Ronneby
SWEDEN

Internet : www.bth.se/tek
Phone : +46 457 38 50 00
Fax : + 46 457 102 45
ABSTRACT

This thesis examines and analyzes the use of data mining techniques and simulations as a forecasting tool.

Decision making process for business can be risky. Corporate decision makers have to make decisions to protect company’s benefit and lower the risk.

In order to evaluate data mining approach on forecasting, a tool, called IFF, was developed for evaluating and simulating forecasts. Specifically data mining techniques’ and simulation’s ability to predict, evaluate and validate Port Industry forecasts is tested. Accuracy is calculated with data mining methods. Finally the probability of user’s and simulation model’s confidentiality is calculated.

The results of the research indicate that data mining approach on forecasting and Monte Carlo method have the capability to forecast on Port industry and, if properly analyzed, can give accurate results for forecasts.

Keywords: Finance, Forecasting, Data Mining, Simulation, Port Systems
“If a man gives no thought about what is distant, he will find sorrow near at hand.”

Confucius
ACKNOWLEDGEMENTS

I wish to express my deep sense of gratitude and indebtedness to my esteemed guide and supervisor Lawrence Henesey, for his keen interest, competent and generous help, and valuable guidance and also for his constant encouragement throughout the period of this study.

I warmly appreciate the active support and cooperation of my family and dear friends throughout the study. I wouldn’t finish this thesis if they were not there for me.
Table of Contents

ABSTRACT .......................................................................................................................... 1
ACKNOWLEDGEMENTS ..................................................................................................... 3
TABLE OF FIGURES .......................................................................................................... 5
INTRODUCTION .................................................................................................................. 6
  1 BACKGROUND ............................................................................................................. 6
     1.1 FINANCIAL FORECASTING ................................................................................. 7
     1.2 PORT INDUSTRY ................................................................................................. 8
  2 PROBLEM DEFINITION ................................................................................................. 9
     2.1 GOALS .................................................................................................................. 10
     2.2 LIMITATIONS ....................................................................................................... 10
     2.3 MOTIVATIONS ....................................................................................................... 10
  3 METHODOLOGY .......................................................................................................... 11
     3.1 LITERATURE REVIEW ....................................................................................... 12
     3.2 FIELD SURVEY .................................................................................................... 12
     3.3 SIMULATION ......................................................................................................... 12
     3.4 RESEARCH QUESTIONS ....................................................................................... 13
  4 THEORETICAL STUDY ............................................................................................... 13
     4.1 DATA MINING ....................................................................................................... 13
         4.1.1 Data Mining Requirements ......................................................................... 13
         4.1.2 Data Mining Tasks ...................................................................................... 14
         4.1.3 Data Mining Process .................................................................................... 15
     4.2 MULTIVARIATE MODELS ..................................................................................... 15
         4.2.1 Monte Carlo Method ..................................................................................... 16
  5 PRACTICAL STUDY ...................................................................................................... 18
     5.1 INVESTMENT AND FINANCIAL FORECASTING TOOL .................................... 19
         5.1.1 Components and Technology ...................................................................... 19
         5.1.2 Working Principle ......................................................................................... 20
  6 CASE STUDY / EXPERIMENT .................................................................................... 21
     6.1 MOFFATT & NICHOL CASE STUDY .................................................................... 21
         6.1.1 Industrial Background ................................................................................... 21
         6.1.2 Results: Data Mining Outputs – (Output Metrics) ......................................... 22
         6.1.3 Results: Statistical Reports ............................................................................ 32
  7 DISCUSSIONS ............................................................................................................. 36
  8 CONCLUSIONS & FUTURE RESEARCH .................................................................... 37
REFERENCES ..................................................................................................................... 38
IFF MANUAL ....................................................................................................................... 41
TABLE OF FIGURES

FIGURE 1 - FRAMEWORK FOR FORECASTING AND PLANNING [8] ................................................................. 8
FIGURE 2 - COMPONENTS OF FORECAST [8] .................................................................................................. 9
FIGURE 3 - TRIANGULATION METHODOLOGY .......................................................................................... 11
FIGURE 4 - DATA AND KNOWLEDGE MINING CYCLE [21] ......................................................................... 15
FIGURE 5 - AN EXAMPLE OF INDIRECT RULES ......................................................................................... 16
FIGURE 6 - CATEGORIZATION OF MULTIVARIATE MODELS AND MONTE CARLO METHOD [8] ............. 16
FIGURE 7 - MONTE CARLO METHOD FORMULA ...................................................................................... 17
FIGURE 8 - OUTPUT FILE SAMPLE ............................................................................................................. 22
FIGURE 9 - STUDY CASE ARFF FILE .......................................................................................................... 23
FIGURE 10 - RESULT OF CONJUNCTIVE RULE ......................................................................................... 24
FIGURE 11 - RESULT OF M5 RULE ............................................................................................................. 25
FIGURE 12 - RESULT OF ZEROR ............................................................................................................... 26
FIGURE 13 - RESULT OF DECISION STUMP ............................................................................................ 27
FIGURE 14 - RESULT OF M5P .................................................................................................................. 28
FIGURE 15 - RESULT OF ADDITIVE REGRESSION ..................................................................................... 29
FIGURE 16 - RESULT OF LINEAR REGRESSION ........................................................................................ 31
FIGURE 17 - FORECAST SUMMARY REPORT PAGE 1 .............................................................................. 32
FIGURE 18 - FORECAST SUMMARY PAGE 4-5 .......................................................................................... 33
FIGURE 19 - SIMULATION REPORT PAGE 1 .............................................................................................. 34
FIGURE 20 - SIMULATION REPORT 2-5 ................................................................................................... 35
FIGURE 21 - COMPARISON OF DATA MINING METHODS ...................................................................... 36


INTRODUCTION

This research will examine and analyze the use of data mining techniques and Monte Carlo method as a forecasting tool. Specifically seven different data mining methods’ ability to predict future values, to find relationships between given values of forecasts and to prove whether foreseen values are probable or not. And generating Monte Carlo method based simulations to be prepared for all possible cases while making decisions. Accuracy will be calculated with the analysis of methods. Finally, the probability of the model’s forecast being correct will be calculated using conditional probabilities in the result of methods. While only briefly discussing data mining techniques and Monte Carlo method, this research will determine the feasibility and practicality of using data mining and Monte Carlo method as a forecasting tool for the individual or business investors.

The contribution of this research is two-fold. First, it provides a methodological contribution to empirical studies on decision making under certainty and further understandings on combining Data Mining, Monte Carlo method and forecasting. Second, it provides a novel forecasting tool in Port industry which validates the offered forecasts, simulates forecasts to analyze the risk probability and helps decision makers with presenting statistical reports, graphics and algorithmic results upon system variables. Since Data Mining approach in Port industry is an undiscovered area of modern science, this renders the research and developed tool on this topic as distinctive.

Thesis is organized in 7 chapters. Chapter 1 describes the backgrounds of the paper, financial forecasting and Port industry. Chapter 2 focuses on defining the problem, goals and limitations of this thesis and the motivations to choose the problem as a thesis project. Methodology and the research questions are in Chapter 3. In Chapter 4, theoretical methods for the project are explained as Data Mining and Monte Carlo. Developed tool and its components are described in Chapter 5. Chapter 6 is focused on Moffatt & Nichol case study and its results. Chapter 7 covers the conclusion of the thesis.

1 BACKGROUND

Over the last 30 years, with the uptrend of information technology related to the port infrastructure increased the potential of the marine and port transportation as well as the progress of the transport industry as a whole. [1] The global marine ports and services market generated total revenues of $45.3 billion in 2007, representing a compound annual growth rate (CAGR) of 7.1% for the period spanning 2003-2007. However the growth rates are expected to drop in over the next five year with the effect of world-wide financial crisis in 2008-2009. [2] Over 90% of the world’s ports are owned or administered by public foundations. Conversely, for some explicit reasons, most important development decisions are taken by the governments. While governments are seeking to increase private sector’s share in marine and port business, developers as well as financiers must analysis the risk, determine the total investment risk and control risk wherever they can. [3]

Since Marine and Port Business has large amount of financial potential, trading on the port market has long been perceived as an important investment which can yield large
amount of profits. [4] Application of financial time series forecasting is more challenging comparing with other modern time series forecasting. Financial time series are no stationary and deterministically chaotic. This means there is no complete information that could be reached from the past actions of financial markets to get the dependency between the future and past price. [5] However there are a lot of researches to correlate past and future data in financial markets, which is mostly focused on Data Mining. Due to the huge amounts of data and need for turning such data into useful information, Data Mining has attracted a great attention in the information industry. The discovery of unexpected or valuable structures in large datasets needs some steps as: Cleaning, Integration, Selection, Transformation, Mining, and Evaluation of the data patterns and Presentation of the Knowledge by the help of useful mined data. [6] Researches in this area are range over a wide field. As a result of these researches, some generalized methods are discovered such as Genetic Algorithms, Neural Networking, Fuzzy Logic and Monte Carlo method. In this paper, Monte Carlo method will be used to generate simulations and data mining will be used to see the relation between past and future data.

Monte Carlo Method is used for iteratively evaluating a model using random sets of numbers with respect to inputs. Monte Carlo method is quite suitable for this project since marine financial system involves more than just a couple of uncertain data. In a Monte Carlo method can cover over thousands of evaluations of the model. By using Monte Carlo method, we are turning the deterministic model into a stochastic model which is an essential component for probability and forecasting cases. In this paper, an online investment and financial forecasting tool with Data Mining and simulation system will be explained and analyzed with the help of some real market based data.

There are challenges and problems of using forecasting tools as well as their advantages. Forecasts empower people since the usage of forecasting tools implies that human can modify variables to be prepared for the future. There will be always blind spots in forecasting. Since completely new stuff which doesn’t have any paradigms, can’t be predictable. [24] There is no way to be completely sure about the future. Regardless of any kind of methods, there will be always a variable which represents uncertainty in models.

Mathematical techniques often begin with some initial assumptions and if these are incorrect, then forecast will reflect these errors. However it would be hard to find the problem as it is the fundamental of forecast model.

1.1 Financial Forecasting

Forecasting is often confused with planning. Forecasting is about what the world will look like, while planning concerns what it should look like. [8] Forecasts are required for two basic reasons: the future is uncertain and the full impact of many decisions taken now is not felt until later. Consequently, accurate predictions of the future improve the efficiency of the decision making process. [7] Financial forecasting is one of the important aspects of management. Knowing or having a close estimation of firm’s financial situation over coming time will help firm to revise its business decisions and strategy for future.
Most of the decisions are made with a prospect to influencing where the situation or company will be in the future. In any part of our lives or our environment such decisions are being made, such as: Families save part of their incomes for their children’s college fund in the future or to bring themselves into safety for their retirement; a company opens new stores or builds new factories to meet the future demand for its products; a broker or an investor in stock market buys some stock in an expect of earning some money in the future, business people buy foreign currency or make the business with foreign currency, when they work internationally, to reduce risk of money loss from exchange rates. [7] These kinds of investments require some prediction of the future behavior of variables so that a decision can be made of what will happen in the future and what kind of precautions can be taken. In financial markets the relations between the present and future values of resources are important for forecasts. The relation between current and future exchange rates, short and long-term interest rates, inflation, governments economic and finance policy, growth of the home and world economy and many more have important implications for planning the future. [7]

Financial forecasting is a time consuming process which requires being patient. The reason is necessary information for assumptions need to be collected carefully for the reliability of the forecast. All variables need to be clarified; for example expected cash demands for the future, profit, staff need and more. In case of uncertainty on these variables, assumptions should be made on business history, or increase in the revenue on expected changes in the market.

Forecasts of the future state of the economy are an essential input into the decision-making process. Once forecast completed, it should be reviewed periodically. An updated forecast can enable you to see if goals in the past reached or not. [5] A correct forecast can be extremely helpful, while making wrong forecast can result in high costs such as extra staff or inventory.

1.2 Port Industry

Marine business is raising its share at transportation and shipments of goods. Ports are increasing their revenues and profits day by day and keep investing for more. There
are almost 600 container ports across the world with a combined handling capacity of more than 380 million TEU. The largest ports, those that can handle in excess of 1 million TEU per annum, account for nearly two thirds of global capacity the global marine ports and services market grew by 7.5% in 2007 to reach a value of $45,261 million. [2]

There are researches have been done and more of them are still in process to find a reliable and accurate way of financial forecasting. Most of these researches focus on theoretical studies while the rest of the studies are focusing on applied technologies about forecasting. However these researches still didn’t reach to the goal exactly.

To survive and compete with the business competitors in the market and to satisfy customer demand, it is vital that ports are managed, and future planned carefully. It is necessary to have the right information and track in the light of forecasts, since it’s the only area that shareholders can control.

![Figure 2 - Components of forecast][8]

2 PROBLEM DEFINITION

The idea of research was aroused from Moffat & Nichol Company’s request to Dr. Lawrence E. Henesey, a tool for financial forecasting on port industry to help analysts on decision making process for company’s future actions and risk assessment. The idea came up last year and the project [28] started last year but the results were not applicable to the market. Because of that this thesis is studied the subject and developed the tool again. The combination of all requests created a tool for finding patterns in port management and also assessing risk degree by running some scenarios with Monte Carlo method.

In today’s intensely competitive market, companies require advance planning. Without precise planning, companies will have to deal with extra costs. However a well modeled forecast can protect the companies against these unnecessary costs. Inaccurate forecasts also harmfully affect on the market reliability and on relationship with
customers, suppliers and partners. So the main question that rises up here is whether or not it is possible to make accurate forecasts continuously.

2.1 Goals

This financial forecasting research is focused on applying some well known Data Mining methods, techniques and simulating forecasts by generating random numbers for specific parameters. The goal of the thesis is obtaining answers for research questions which will be mentioned next chapter and fulfilling the following objectives:

- Designing and implementing simple and functional open source model and tool
- Processing the vast amount of data and transforming those data into information
- Using Data Mining techniques to find out the hidden relationships between variables
- Verifying results with models
- Identifying the comparison of different Data Mining methods
- Identifying risk probability of the project

2.2 Limitations

One of the limitations in this research was the ability of data. System consists of over thousand variables. Some of these variables involve classified data which is hard to obtain from companies as a study case.

Another limitation factor is the lack of papers or other kind of sources on this very specific topic. Forecasting and Data Mining are quite new by meaning of uncovered topics.

2.3 Motivations

In this section, motivations for choosing this problem as a thesis project are explained.

Computer scientists work mostly as theorists, researchers, or inventors and apply their higher level of theoretical knowledge and innovation to complex problems. New programming tools, knowledge-based systems and systems that enable computers to perform their many applications are taking place on market. The results of these studies, technological impossibilities are vanishing every day. In this highly interactive environment, business interests started to install new technologies into their decision process.

Forecasting involves complex models and algorithms to compare thousands of variables, which is becoming more possible. Professionals for forecasting and analyzing the results are needed in the sector to be prepared for the future. As Tomi T. Ahonen, who is a well-known strategy consultant, says that "The value of predicting the future and forecasting is not in knowing exactly what the future is going to be, it is rather to be
prepared for upcoming. Industry can’t wait for the unknown future while their competitors are getting prepared for that.”

In the big picture, the project involves Financial Forecasting, Data Mining and Port Industry. Investment and Financial Forecasting has a big potential as it is explained above. Data Mining is still a new field on statistics and information technology by reason of unexplored areas. And the last but not least one, Port Industry, there are huge amounts of money is involved and still keep growing, as it is explained in section 1.2. When all these components come together, the career possibilities can be seen easily which makes this project challenging and interesting.

3 METHODOLOGY

Methodology provides tools and techniques that researchers can use for gaining knowledge, firmer understanding and solving problem.[17] Researcher’s analysis of methods, rules, and postulates applied in the thesis is covered by the meaning of methodology. Parallel with the complexity of the research, researcher may decide for using more than one methodology in the study.[15] As a result, both quantitative and qualitative approach are used in this thesis by triangulation methodology. Triangulation methodology is most appropriate for evaluations that seek to answer questions concerning the quality, implementation, and outcome. [16]

The most common methods used in today’s researches are action research, literature view, survey and case study / experiment. [18] In action research, the main goal is to find out how to do things in a better way, instead of knowing what is wrong. [19]

The methodology used while conducting the research is as follows:

- Conduct a literature search of books, magazines, journals, articles and the World Wide Web on the topic of data mining techniques on Port industry and Monte Carlo method.
- Identify the theory behind the Data Mining and Monte Carlo method.
- Determine what to forecast in the system and the future point for the forecast (5 years into the future).
- Simplify all the variables in port business by determine the inputs to data mining and Monte Carlo method will use.
- Develop the forecasting tool by considering both theories
- Attempt to forecast using multiple data mining techniques.
- Compare data mining technique results.
- Determine probabilities whether forecast is successful using Data Mining
- Attempt to forecast using randomly assigned values of Monte Carlo method

### 3.1 Literature Review

Review of journals, periodicals, and other research publications related to the subject area was executed during the initial phase of the research and updated throughout the research. Several research papers, journals, conference papers and books needed to be reviewed to understand the operation style of ports and forecasting. Even though, incredible amount of databases are searched under the supervision of Dr. Lawrence E. Henesey, a direct paper, research or other material was not found on this topic since this topic is very specific and new. Other than that the papers and other material about Data Mining, Monte Carlo and Port Industry are covered in this thesis.

### 3.2 Field Survey

During the process of the thesis, a field trip was done to the Port of Karlshamn, Sweden. The trip allowed us to understand more about port operations and establish good even if short relationships with the employees and the end users, by collecting and gathering information at the local level as field survey.

This step enhanced researcher’s understanding of port operations. The results and acquisitions from this step of methodology helped and used in the next step of the study, which is case study simulation.

### 3.3 Simulation

Simulation is a means of conducting experiments of system behavior on models mimicking the real system with sufficient accuracy. [17] Simulated experiments accelerate ‘wait and see’ concerns about the future real systems.

Generated simulation in this project used as a way to understand general system dynamics and generates insights. The system has functions that perform operations with a mixture of controllable system parameters and uncertain variables.
The tool which is developed in this project is called *IFF*. Simulation is handled by *IFF* and Excel. Monte Carlo method is used for simulating next 5 or 10 years and repeats forecast 3000 times to get more accurate results.

### 3.4 Research Questions

The following research questions allow the research to meet the objectives proposed:

- RQ1: How can Data Mining be useful in a forecasting tool?
- RQ2: How can we analyze, evaluate and structure vast amount of data by using Data Mining techniques?
- RQ3: How can we apply financial or forecasting models with Data Mining techniques together?
- RQ4: Is using Monte Carlo method provide helpful accurate results?

### 4 THEORETICAL STUDY

#### 4.1 Data Mining

The process of analyzing data from different perspectives, categorizing it, summarizing into useful information and finding correlations between variables is called Data Mining. [21] On his book [32] David J. Hand, who is the head of the Statistics Section at Imperial College, said that “Data Mining (consists in) the discovery of interesting, unexpected, or valuable structures in large data sets”. Data Mining is a powerful new technology to help companies focus on important information in their resources, which gives them complete control of their own resources. Answers of the time consuming questions can be resolved in a matter of time and with some predictive information results that expert may miss because of low probability. This information can be either automated prediction of trends and behaviors by examining past data or automated discovery of unknown patterns. [20] Data Mining can be helpful in any kind of business sector, such as marketing, banks, law, researches and many more.

#### 4.1.1 Data Mining Requirements

Investors and Data Mining Analysts needs to be careful during the design of the Data Mining System. On the ground that project should meet the some requirements to be able to use data mining effectively and to get proper results.

In general outline, there are 5 requirements for Data Mining. [31] These are:

1. Large amount of correct data which is placed in a relational or multidimensional database. If the accuracy of data decreases, as well as the reliability of the predictions will decrease.
2. Data Mining Algorithm
3. Data model which algorithms work on should be well-designed.
4. Multiprocessor-based computer which has an access to database. Large databases can result with very complex models which need to be processed on powerful hardware.
5. Someone who knows what algorithms and models are doing
It can be understood better with an example [27] by showing how data mining is used in some banks to help loan decisions and prevent fraud attempts. The steps of decision process are as follows:

1. Get the previous customer records such as age, sex, marital status, occupation, number of children and etc.
2. Examine previous customer data with an algorithm that identifies characteristics of customers who took a specific kind of loan and who did not.
3. Eventually, some rules will be appeared to be a good candidate for that loan
4. These rules are then used to identify such customers on the database
5. Next, another algorithm is used to sort the database into groups of people with many similar attributes, with the expectation of revealing hidden, interesting and unusual patterns.
6. Then patterns revealed by these customers are interpreted by the data miners, in collaboration with bank personnel
7. Finally, the decision for the customers will be made with the help of risk degree and etc. of the customer according to revealed patterns.

Data Mining techniques are not exactly %100 precise about the prediction since it needs to be examined by professional data miners and business people; it does help the banks reduce their losses by preventing fraud attempts.

4.1.2 Data Mining Tasks

In Figure 4 - Data and Knowledge Mining Cycle [21], data and knowledge mining cycle is represented. These analysis steps focus on a different aspect or task. [32] proposes the following data mining tasks for data mining.

A. Classification

Classification technique analyzes data into predefined categories. It helps to keep organized and also quick accessibility to sources. Tagging or labeling, which is used in today’s internet world, is a result of classification. Categorizing people with gender and age range can be an example of classification.

B. Clustering

The application principle is quite similar with Classification. The only difference is Clustering analyzes data and categorizes them together instead of predefined classes. The algorithm of clustering tries to put similar data together. Clustering helps to find out hidden predictive data.

C. Regression

Regression technique tries to model the data with a function. Simplest example for this is ‘y = Ax + B’, algorithm determines appropriate values for ‘A’ and ‘B’ to find out ‘y’. Genetic Programming is a good example for regression. Data is represented as tree structure which helps to be easily evaluated recursively.

D. Association Rules Learning

Pattern mining is a Data Mining technique that searches for existing patterns or relationship between data, these patterns might be regarded as small signals in a large
ocean of noise. In this context patterns often means association rules. By analyzing supermarket data or examining customers behavior in terms of purchased products.[22] For example, if a manager knows by pattern mining that 70% of customers who purchase wine also purchase cheese, this could lead to a marketing decision by locating wine section and cheese section closer. [21]

4.1.3 Data Mining Process

Even though there are different Data Mining techniques and methods, the process for mining is same with all of them. These 5 steps are:

1. Selection: Objective variables and also dependent variables are selected as target data at the beginning of the process.
2. Preprocess: Data properties should be analyzed at this step.
3. Transformation: Depending on the step 2, different kinds of techniques should be applied so that the data can be ready for Data Mining.
4. Data Mining: In this step most suitable tasks applied to data.
5. Evaluation: Results of Data Mining tasks compare to each other and evaluate.

Figure 4 - Data and Knowledge Mining Cycle [21]

4.2 Multivariate Models

Multivariate model is a kind of complex "What if?" scenarios. In multivariate model, impact of a factor in the forecast can be ascertained by manipulating the values. Multivariate model is useful when multiple factors affect the outcome. In these cases, using one variable based model can mislead the studies by ignoring significant relationships, while multivariable model can show the relation. For example;
These models are applied in several business areas such as finance, banking, insurance etc. The best-known multivariate models are used to assess the stocks, to help analysts find out the path of the real values and also financial advisors or analysts use multivariate models to determine the risk and estimate the cash flow. [12]

4.2.1 Monte Carlo Method

The name “Monte Carlo” was aroused by John von Neumann, Nicholas Metropolis and Stanislaw Ulam in the late 40s. During the World War II, these three scientists were studying on neutron diffusion in a nuclear-atomic bomb project called Manhattan Project. Ulam had personal interest in poker; his uncle was going Monte Carlo each year for gambling. After he noticed the similarity of statistical simulations and chance games, they named the method as Monte Carlo in [10]. [9] Monte Carlo is now used routinely in many diverse fields, from the simulation of radiation transport in the earth's atmosphere to the simulation of a Bingo game.

Monte Carlo Method is a method of analysis based on artificially recreating a chance process, running it many times, and directly observing the results. [13] Monte Carlo simulation is often used in business for decision and risk analysis. [14] In Monte Carlo simulation, a deterministic model is evaluated several times by using random numbers as inputs.

[Figure 5 - An example of indirect rules]

[Figure 6 - Categorization of Multivariate Models and Monte Carlo Method]
This method is used when the model has more than one uncertain variable or nonlinear. Monte Carlo method is applied as simulation which is an artificial model of a real system in a computer environment to analyze and understand the system. [29] Main reason for using Monte Carlo simulation is to study different statistics computed from a bunch of sample data. By using Monte Carlo simulation, system is also tested for extreme cases as a risk analyzer. The procedure will be as specifying the values of important parameters in the system and randomizing the other variables. A typical Monte Carlo simulation can involve over 1000 evaluations of the model.

Figure 7 - Monte Carlo method formula

In Figure 7 - Monte Carlo method formula, essential Monte Carlo method formula is shown, this formula applied by following the steps: [13]

1. Create a parametric model as \( y = f(x_1, x_2, \ldots, x_q) \)
2. Generate random inputs, \( x_1, x_2, \ldots, x_q \)
3. Evaluate the model and store the results as \( y_i \)
4. Repeat steps 2 and 3 for \( i = 1 \) to \( n \)
5. Analyze the results using statistics, histograms, intervals

Uncertainty and Risk is two important conceptions for Monte Carlo method. Uncertainty is an objective notion over the situations which cannot be controlled such as weather or tomorrow’s Euro (€) – Dollar ($) parity. On the other hand, Risk is subjective which differs on everybody and situation. A Factory owner buys necessary raw material for his product in US Dollars but he sells his product in Euro. Now for this person, risk is Euro’s increasing value against US Dollars. However if he would buy raw material in Euro and sell his product in US Dollars, risk would be Euro’s value loss against US Dollars. So risk involves a formula with uncertain variables. In this point, to get a risk, uncertain variables need to be initialized. If these uncertain variables replaced with average values, it will just give an average result not risk percentage because risk is never about averages. In his paper [33] Savage summarizes with “Plans based on average assumptions will be wrong on average” and presents flaw of averaging as Seven Deadly Sins of Averaging:

1. The Family with 1½ children: Average scenario like the family with 1½ children is not existing. For example, a bank has 2 main portfolio types, students with $10,000 income and young businessmen with $70,000 income. If bank decides to build products or services for customers with the average income of $40,000, it would be just disaster and non-sense.

2. Why Everything is Behind Schedule: A project with 10 separate sections in a parallel development. The timeline for each section is uncertain and independent, but in average it is three months. It would be tempting to approximate timeline as
three months for the entire project but chance of this to happen is same as flipping 10 sequential tails with a fair coin.

3. The Egg Basket: Consider putting 10 eggs all in the same basket, versus one by one in separate baskets. In the first one there is a possibility of %10 for losing all the eggs, however there is only one chance in 10 billion of losing all the eggs in second one.

4. The Risk of Ranking: It is common to rank investments from best to worst, and then customer starts to invest from top to down on the list. In accordance with ranking, fire insurance is not good since on average it loses money. However it would change everything if customer has a house or a real estate in the portfolio.

5. Ignoring Restrictions: A product provides capacity equal to the average of uncertain future demand. It is common to assume that profit with average demand will be average. However it doesn’t work like that. If actual demand is less than average, profit will drop. But if actual demand is greater than average, profit will be up with a restriction of capacity. At the end there is a bad scenario without any limits and good scenario with a restricted limit, this will lead less profit than average profit.

6. Ignoring Options: A product factory with known marginal production costs and an uncertain future product price. Averaging will lead to an average value for the factory. If product prices go above average, factory’s value will go higher. But if product prices go below the marginal cost, there is an option to halt production instead of losing money. This option makes the company worth more than average value with respect to average product price.

7. The Double Whammy: A firm which sells perishable goods with uncertain demand with an average demand based stocked. In an average case there is no cost for managing goods. However, if demand is less than average then there will be extra cost to get rid of bad goods and if demand is greater than average there will be lost sales costs since there is a limited amount of goods. So average demand’s cost is zero, but average cost is positive.

These rules show another point of view to look at average valued cases which may lead to disasters. [33]

To sum up, Monte Carlo analyzes are not just for finance people but also for many other businesses. It helps for decision making, by proving that every decision have some effect on the result and creating a picture of risk. Every investment, firm, people has different risk tolerances, as such, risk model of everything needs to be calculated separately and carefully.

5 PRACTICAL STUDY

During the process of thesis, a tool is developed to generate forecasts and simulation for risk analysis. This tool is developed within the outlines of another research group [28] of Dr. Lawrence Henesey. The tool is open source and cross platform. The only thing needs to be done is installing the tool and its requirements to the server. The next step is using the tool online from any computer without any restrictions.
Basically, the features of the tool is
- Getting inputs from users
- If users have their own forecasts, applying data mining techniques on these forecasts and analyzing the risk probability
- If user doesn’t have their own forecasts, generating a simulation and running this simulation thousands time.
- Applying data mining techniques to generated simulation, so that risk probability for simulation can be analyzed.
- Presenting the results of data mining techniques and statistical reports to the users for their forecasts and simulations.

5.1 Investment and Financial Forecasting Tool

5.1.1 Components and Technology

The user-interface is written in PHP and JavaScript, Java for Data Mining techniques, VBA for automation of Monte Carlo functions and simulation. The purpose of this Investment and Financial Forecasting Tool is to generate forecasts or simulations for a period interval. The result is a set of reports and results that satisfies these requirements and minimizes the cost of generating forecasts. All together, these results and reports are good start for port management.

A. XAMPP

XAMPP [26] is a free and open source cross-platform web server package, consisting mainly of the Apache HTTP Server, MySQL database, and interpreters for scripts written in the PHP and Perl programming languages.

The reasons for XAMPP is chosen in this project are:
- It is easy to install, set up and use
- It contains a number of useful packages that make accelerate PHP content.
- It is cross platform. It has been thoroughly tested on the SUSE, Red Hat, Mandrake, and Debian Linux distributions, as well as on Windows® and Solaris.

The important programs for this project inside XAMPP’s this version are:
- Apache 2.2.12 (IPv6 enabled) + OpenSSL 0.9.8k
- MySQL 5.1.37 + PBXT engine
- PHP 5.3.0
- phpMyAdmin 3.2.0.1

The roles of XAMPP in the project are:
- Creating a server environment for php files to work properly
- Providing database system for user and file management

XAMPP v1.7.1 is included in the project setup file.

B. CutePDF™ Writer & Converter

CutePDF™ Writer is a free plug-in application which installs itself as a "printer subsystem". This enables virtually any Windows applications to create PDF documents with normal print function. [25]
The role of CutePDF™ Writer & Converter in the project is:
- Transforming necessary computed statistical excel files into pdf files since pdf is a standard for secure and reliable distribution of documents

CutePDF™ Writer v2.7 and Converter (GPL Ghostscript) v8.15 are included in the project setup file.

C. WEKA

WEKA (Waikato Environment for Knowledge Analysis) is open source software which includes collection of machine learning algorithms for data analyzing and predictive (forecast) modeling. The algorithms can be applied to datasets.

One of the important stones of the tool is Data Mining for given variables. To support this effort, Weka has been chosen which not only gives easy access to a variety of machine learning techniques through an interactive interface, but also incorporates those pre- and post-processing tools that we have found to be essential when working with real-world data sets. [23]

During the process of the studies, Weka was used from their online service, however 2 weeks before the submission Weka Online service is shut down. For the sake of the project, our tool has been modified to work with Weka program as setup. Weka Online can be added to program in the future depending on Weka Online’s situation at that time. The reason for choosing Weka Online is mostly about mobility of the tool, other than that Weka Online and Weka setup is working same in every aspect.

The role of Weka in the project is:
- Transforming the data from User Forecast or Scenario Simulations into usable knowledge by applying necessary Data Mining methods
- Finding out relationships between variables and giving proper output
Weka 3.6.1 is included in the project setup file.

D. Excel

Microsoft Excel is a spread sheet application. It features calculation, graphing tools, pivot tables and a macro programming language called VBA. Microsoft Excel is common in business finance due to wide range of functions and easy to be modeled.

The role of Excel in the project is:
- Applying models
- Running simulation with Monte Carlo functions
- Running macro code on user files which are also excel files
- Being middle layer between calculated statistics and pdf files

Microsoft Office Excel 2007 is distributed by Microsoft. User needs to install a licensed version of this product.

5.1.2 Working Principle

IFF consists of two parts as Forecast Statistics and Simulation. In the first part, user needs to:

1. Fill in the necessary information into our system’s template and upload it
2. Assign a name for the user forecast file
3. Activate forecast models and produce statistics
4. Save the excel file for further use of data mining techniques
5. Print Statistics
6. Apply data mining techniques to user scenarios or output metrics from user’s statistics.
7. Analyze the results of data mining techniques in the meaning of Correlation Coefficient to find out the relationships and risk probability.

In the second part of the IFF, system automatically runs necessary macros and Monte Carlo functions.

1. Run functions in MonteCarloSimulation.xls
2. Calculate results
3. Repeat the step from step 1 till the number of forecasts reaches up to 3000
4. Analyze the results and calculate profit of the system
5. Analyze the profits and determine the best case with macro functions
6. Print out Analysis
7. Save the necessary information from MonteCarloSimulation.xls into new generated xls file
8. Apply data mining techniques on generated xls file
9. Analyze the results for risk probability

6 CASE STUDY / EXPERIMENT

There are 2 case studies for this paper, which are based on statistics of Moffatt&Nichol Co. and YILPORT [11] Container Terminal and Port Operators Inc. Due to YILPORT Company’s internal bureaucracy, the statistics and the results for this case study will be added to the paper in revision process.

6.1 Moffatt & Nichol Case Study

This study is based on the values of an international Marine Engineering company, having several types of services on port and marine industry.

6.1.1 Industrial Background

Moffatt & Nichol was founded in 1945 to provide design engineering services to the evolving maritime infrastructure on the west coast of the United States. Firm earned a reputation for innovation and creativity after the success in coast design and port planning. By the next decade, Moffatt & Nichol had expanded its services to support the larger demands of the goods movement industry in the United States, Canada, the United Kingdom, and Latin America. Today the firm provides clients worldwide with customized service.

Employees: 500 professionals worldwide

Business Segments:
- Coastal, environmental and water resources
- Ports, harbors and marine terminals
- Rail and transportation systems
Urban waterfronts and marinas

Services:
- Financial and economic modeling
- Market forecasting
- Feasibility studies
- Business planning
- Master planning
- Engineering and design
- Program management
- Project management
- Construction management

Clients: A variety of public and private entities worldwide

Headquarters: Moffatt & Nichol is headquartered in Long Beach, California, and operates from offices throughout North America, Europe, Latin America and the Pacific Rim.

The necessary information for Moffatt & Nichol is taken from company’s website and their fact sheet. [34]

6.1.2 Results: Data Mining Outputs – (Output Metrics)

In this section results of the study case will be explained for each data mining technique. All the results consist of three parts:

```
Options: -M 3 -M 2.0 -P -1

Single conjunctive rule learner:

|averageRatePerMove <= 152.5) -> Profit = 10124.85 |

Time taken to build model: 0.02 seconds
Time taken to test model on training data: 0.02 seconds

--- Error on training data ---

Correlation coefficient 0.8924
Mean absolute error 4196.696
Root mean squared error 4826.8575
Relative absolute error 50.0753 %
Root relative squared error 45.1783 %
Total Number of Instances 15

--- Cross-validation ---

Correlation coefficient 0.8953
Mean absolute error 8844.0843
Root mean squared error 10430.9267
Relative absolute error 106.5513 %
Root relative squared error 95.4397 %
Total Number of Instances 15
```

Figure 8 - Output file sample
Part 1 is result, part 2 and 3 are validation of the techniques. Part 1 is written in different format for each data mining technique because of different outcomes. In this part information is for decision makers since found rules, models, expected profit are represented in here. These are the results which user is looking for to make predictions. It needs to be examined well.

Part 2 and 3 are formatted same in every function and help the user about how much these functions are reliable on these results. Correlation coefficient shows the accuracy or relationship probability of the variables. The best case for correlation coefficient is %100 or 1 which implies that there is a complete relation between the variables. User will define their lower limit themselves for this attribute. Relative absolute error and Root relative squared error indicates whether the algorithm is valid for these inputs or not. If the results are close or greater than %100 that means the scheme does not do any better than just calculating the mean of target values. So the best case for Relative absolute error and Root relative squared error would be close to %0.

In case of huge differences between Part 2 and 3, the stratified Cross-validation (Part3) paints a more realistic picture. The results of this study case will be examined in next pages by assuming the accuracy limit is %75.

**A. ARFF File**

ARFF is basically a CSV format with some extra headers to specify what type each attribute is. It is very easy to add some headers and convert a CSV file into an ARFF file.

![Study Case ARFF File](image-url)

Figure 9 - Study Case ARFF File

All header commands start with `@` and all comment lines start with `%`. Comment and blank lines are ignored. Line 1 is a comment line. Line 2 is a header command that names the dataset; in this case the data set is called ‘userExcel’.
Lines starting with ‘@attribute’ define all attributes followed by the name and the type of attribute. There are two main types of attributes, numeric and nominal. Numeric attributes are defined as either ‘real’, ‘integer’ or just ‘numeric’. Nominal attributes are defined by placing in brackets all the possible values an attribute can take. [30]

In the cases like this there are many class variables, as is the case with association rules where our tool will test how well each attribute can be predicted based on the other attributes.

B. Conjunctive Rule

Singe conjunctive rule learner predict for numeric and nominal class labels.

Options: -N 3 -M 2.0 -P -1

Single conjunctive rule learner:
--------------------
(revenuesThroughputTEUs > 559000) => PROFIT = 39206.705

Time taken to build model: 0.02 seconds
Time taken to test model on training data: 0.02 seconds

=== Error on training data ===

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient</td>
<td>0.7652</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>4699.003</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>8387.2995</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>56.0689 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>78.5032 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>

=== Cross-validation ===

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient</td>
<td>-0.5978</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>9183.6353</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>12885.0835</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>87.6639 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>96.4812 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 10 - Result of Conjunctive Rule

The assessment is:

- **Step 1 – Result:** If the revenue through put by TEU is more than 559.000$, Profit will be 39.206,75$. 
Step 2 – Accuracy: Correlation coefficient 1 (CC1) is 76 and CC2 is -60 which doesn’t look good. CC2 indicates that there is an indirect relation but still not enough.

Step 3 – Validation: Relative absolute error 1 (RAR1) is 56, Root relative squared error 1 (RRSR1) is 78, RAR2 is 89 and RRSR2 is 96. As it is seen from the RAR and RRSR values are too close to 100 which means this function is not valid for the inputs since it is just giving mean value of results.

C. M5 Rule

M5 pruned model converts k-valued nominal attributes into k-1 binary attributes using the method and generates a rule to be used later in M5P tree. M5 Rule is well-known technique in financial prediction industry.

```plaintext
Options: -M 4.0
M5 pruned model rules
(using smoothed linear models) :
Number of Rules : 1
Rule: 1

PROFIT =
  0.1127 * revenuesThroughputTEUs
  - 28970.0719 [5/11%]

Time taken to build model: 0.09 seconds
Time taken to test model on training data: 0.02 seconds

=== Error on training data ===

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0.9939</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>1044.8031</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>1175.1892</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>12.4667 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>10.9995 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>

=== Cross-validation ===

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0.8163</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>6187.439</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>8116.1281</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>59.0632 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>60.7721 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>
```

Figure 11 - Result of M5 Rule
The assessment is:

- **Step 1 – Result:**
  - \( \text{PROFIT} = 0.1127 \times \text{revenuesThroughputTEUs} - 28970.0719 \ [5/11\%] \)
  - Profit is depending on revenue through put by TEU in the meaning of given rule.

- **Step 2 – Accuracy:** Correlation coefficient 1 (CC1) is %99 and CC2 is %81 which is quite good. Both of them shows that there is a high accuracy for found profit function.

- **Step 3 – Validation:** Relative absolute error 1 (RAR1) is %12, Root relative squared error 1 (RRSR1) is %10, RAR2 is %59 and RRSR2 is %60. As it is seen from the RAR1 and RRSR1 values are close to %0 while RAR2 and RRSR2 is about %60. Base on these results, there is a high accuracy potential and even though Cross-validation ratio is not perfect it has high potential too when it is combined with RAR1 and RRSR1.

### D. ZeroR

ZeroR provides a useful indication of the system’s worst performance, since a non-random prediction scheme does better. And ZeroR is a baseline classifier that simply identifies the class that is mostly finding in the dataset, and predicts all variables to be in that class. For example, there is a community (dataset) and in this community has people (variables). If %60 of these people is old and %40 is young, ZeroR will treat (classify and apply) to people (variables) as everybody is old since they are majority.

**ZeroR predicts class value: 29193.638**

Time taken to build model: 0.02 seconds
Time taken to test model on training data: 0.02 seconds

*** Error on training data ***

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>8380.7704</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>10664.0212</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>100 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>100 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>

*** Cross-validation ***

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>10475.963</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>13355.0265</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>100 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>100 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 12 - Result of ZeroR
The assessment is:

- **Step 1 – Result:** ZeroR predicts class value: 29193.638
- **Step 2 – Accuracy:** Correlation coefficient 1 (CC1) is %0 and CC2 is -%100 which is bad.
- **Step 3 – Validation:** All the necessary values are %100 which means this function is not trustworthy at all.

**E. Decision Stump**

This is a very simple one-level decision tree. Each decision stump consists of a single decision node and two prediction leaves.

---

**Figure 13 - Result of Decision Stump**

<table>
<thead>
<tr>
<th>Decision Stump</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifications</td>
</tr>
<tr>
<td>revenuesMoves &lt;= 250000.0 : 10124.85</td>
</tr>
<tr>
<td>revenuesMoves &gt; 250000.0 : 33960.835</td>
</tr>
<tr>
<td>revenuesMoves is missing : 29193.638</td>
</tr>
</tbody>
</table>

Time taken to build model: 0.02 seconds
Time taken to test model on training data: 0.02 seconds

--- Error on training data ---

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0.8924</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>4196.696</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>4821.1658</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>50.0753 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>45.125 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>
The assessment is:

- **Step 1 – Result:**
  - revenuesMoves <= 250000.0 : 10124.85
  - revenuesMoves > 250000.0 : 33960.835
  - revenuesMoves is missing : 29193.638
  
  If the revenueMoves is smaller than or equal to 250,000 Profit will be 10,124.85$.
  
  Else if revenueMoves is greater than 250,000 Profit will be 33,960.835$.
  
  In case of missing revenueMoves, Profit will be 29,193.638$.

- **Step 2 – Accuracy:** Correlation coefficient 1 (CC1) is 89% and CC2 is 22% which is in between.

- **Step 3 – Validation:** Relative absolute error 1 (RAR1) is 50%, Root relative squared error 1 (RRSR1) is 45%, RAR2 is 104% and RRSR2 is 92%. As it is seen from the RAR2 and RRSR2 values are close to 100 which means this function is not trustworthy.

**F. M5P**

M5 Pruned Model tree applies the rule found in M5 Rule.

![Figure 14 - Result of M5P](image)

**Options:** -H 4.0

M5 pruned model tree:
(learning smoothed linear models)

LN1 (5/11x)

LN num: 1

Profit =
0.1127 * revenuesThroughputTEUs - 28970.0719

Number of Rules : 1

Time taken to build model: 0.09 seconds

Time taken to test model on training data: 0 seconds

*** Error on training data ***

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0.5939</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>1044.8031</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>1175.1032</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>12.4667 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>10.5985 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>

*** Cross-validation ***

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>0.5163</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>6187.435</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>8115.1281</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>59.0632 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>60.7721 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>5</td>
</tr>
</tbody>
</table>
The assessment is:

- **Step 1 – Result:**
  LM1 (5/11%) LM num: 1
  \[ \text{PROFIT} = 0.1127 \times \text{revenuesThroughputTEUs} - 28970.0719 \]
  If the revenue through put by TEU is more than 559.000$, Profit will be 39.206,755.

- **Step 2 – Accuracy:** Correlation coefficient 1 (CC1) is %99 and CC2 is -%81 which shows the accuracy clearly.

- **Step 3 – Validation:** Relative absolute error 1 (RARE1) is %12, Root relative squared error 1 (RRSRE1) is %59 and RRSRE2 is %60. Both validation results are not bad so this function is valid.

**G. Additive Regression**

Figure 15 - Result of Additive Regression
The assessment is:

- **Step 1 – Result:** Additive Regression gave 10 models with decision stump. At the end of this rule, we have 10 basic relationship based decision trees.
- **Step 2 – Accuracy:** Correlation coefficient 1 (CC1) is %100 and CC2 is %69 which seems good.
- **Step 3 – Validation:** Relative absolute error 1 (RAR1) is %1, Root relative squared error 1 (RRSR1) is %1, RAR2 is %55 and RRSR2 is %66. When both of the parts are considered, the results seem accurate and valid with high correlation and average error.

**H. Linear Regression**

Options: -8 0 -R 1.0E-8

**Linear Regression Model**

\[
\text{PROFIT} = 0.004 \times \text{revenuesThroughputTEUs} + 0.0066 \times \text{revenuesMoves} + 43.997 \times \text{revenuesAverageRatePerMove} + 27.4981 \times \text{revenuesOtherRevenuePerMove} + 0.0345 \times \text{revenueTotalContainerRevenue} + 0.0213 \times \text{revenueTotalRevenue} + 0.1357 \times \text{expensesVesselLaborExpenses} + 0.2217 \times \text{expensesYardLaborExpenses} + 0.0381 \times \text{expensesTotalLaborExpenses} + 0.7036 \times \text{expensesMarExpenses} + 0.1451 \times \text{expensesSG&A_includingRent} + 0.0515 \times \text{expensesTotalExpenses} + 0.0351 \times \text{expensesEBITDA} + 7.6812 \times \text{vesselCalls} + 461.7751 \times \text{ShipCranes} + 420.0291 \times \text{TopPick} + 185.6584 \times \text{AverageHourlyWagePerHour_vessel} + 185.6584 \times \text{AverageHourlyWagePerHour_yard} + -102.7528 \times \text{Yard-GatesLaborCostPerMove} + 50.4598 \times \text{AverageUsedTotal} + 0.2836 \times \text{groundSlotsNeededDTG} + 0.945 \times \text{groundSlotsNeededTopPick} + 0.3943 \times \text{groundSlotsNeededSidePick} + 0.5972 \times \text{revenueTeusPerVesselCall} + 0.0079 \times \text{revenueTeusPerBerth} + -0.0172 \times \text{revenueTeusPerShipCranes} + 1.0257 \times \text{revenueMovesPerVesselCall} + 0.0136 \times \text{revenueMovesPerVesselBerth} + -0.0296 \times \text{revenueMovesPerVesselShipCranes} + 0.9719 \times \text{vesselOperationsTotalMovesPerCall} + 963.5502 \times \text{vesselOperationsGangs} + -90.2396 \times \text{vesselOperationsHoursBerth} + -462.2935 \times \text{vesselOperationsBerthUtilizationRate} + 185.6584 \times \text{AdjustedWagePerHour_vessel&yard} + 70.0072 \times \text{yardOperationsGateHours} + -136.6076 \times \text{yardOperationsGateProcessingTime(minutes/transaction)} + -8.6952 \times \text{yardOperationsLanesHoursRequired} + 185.6584 \times \text{adjustedWagePerHour} + 5882.6017 \times \text{capacityUtilization} + -95801.593
\]
Figure 16 - Result of Linear Regression

The assessment is:

- Step 1 – Result: A Profit equation which consists of several variables with coefficients.
- Step 2 – Accuracy: The best accuracy result for this case study is Linear Regression. Correlation coefficient 1 (CC1) is %100 and CC2 is %96 which is a highly correlated variables are found.
- Step 3 – Validation: Error statistics is also quite low comparing with other functions. Relative absolute error 1 (RAR1) is %0, Root relative squared error 1 (RRSR1) is %0, RAR2 is %29 and RRSR2 is %39. It shows that the equation is trustworthy.
6.1.3 Results: Statistical Reports
6.1.3.1 Forecast Statistical Summary Report

In this part, generated report’s first 3 pages contain business terms such as TEU Growth, Rate per Box, Total Expenses/Percentage revenues, Profit and EBITDA. User already knows these terms since he/she gave the values at the beginning. These are revenues, expenses, operation metrics and most importantly Profit.

<table>
<thead>
<tr>
<th>Summary Report</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenues</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Throughput (TEUs)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>Throughput Moves</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>Average Rate per Move</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>Other Revenue per Move</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>Total Container Revenue ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>Total Revenues</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vessel Labor Expenses ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>% of Revenues</td>
</tr>
<tr>
<td>Yard Labor Expenses ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>% of Revenues</td>
</tr>
<tr>
<td>Total Labor Expenses ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>% of Revenues</td>
</tr>
<tr>
<td>M&amp;R Expenses ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>% of Revenues</td>
</tr>
<tr>
<td>SG&amp;A (including Rent) ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>% of Revenues</td>
</tr>
<tr>
<td>Total Expenses ('000)</td>
</tr>
<tr>
<td>% Growth</td>
</tr>
<tr>
<td>% of Revenues</td>
</tr>
<tr>
<td>EBITDA ('000)</td>
</tr>
<tr>
<td>% Margin per Move</td>
</tr>
<tr>
<td>Capex</td>
</tr>
<tr>
<td>Maintenance</td>
</tr>
<tr>
<td>Cumulative</td>
</tr>
<tr>
<td>Expansion</td>
</tr>
<tr>
<td>Cumulative</td>
</tr>
<tr>
<td>Profit ('000)</td>
</tr>
</tbody>
</table>

Figure 17 - Forecast Summary Report Page 1
a. EBITDA ($) – Margin (%) graph per year.

b. Throughput(TEU) – Growth (%) graph per year

c. Average Rater Per Move – Growth (%) graph per year

d. Total Expenses($) - % of Revenues graph per year

e. Profit distribution per year

Figure 18 - Forecast Summary Page 4-5
6.1.3.2 Simulation Report

Simulation report is done after 3000 repetition of random variables and consists of statistical and predictive results. Statistical numbers after first page is best case’s values.

![Simulation Report](image)

**Figure 19 - Simulation Report Page 1**
<table>
<thead>
<tr>
<th>Monte Carlo Simulation (Summary Report)</th>
<th>23.08.2020 15:51</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenues</strong></td>
<td></td>
</tr>
<tr>
<td>Throughput (TEUs)</td>
<td>2,869,924.97</td>
</tr>
<tr>
<td>% Growth</td>
<td>13.87%</td>
</tr>
<tr>
<td>Throughput Moves</td>
<td>1,638,857.50</td>
</tr>
<tr>
<td>% Growth</td>
<td>21.49%</td>
</tr>
<tr>
<td>Average Rate per Move</td>
<td>153.34</td>
</tr>
<tr>
<td>% Growth</td>
<td>38.19%</td>
</tr>
<tr>
<td>Other Revenue per Move</td>
<td>159.36</td>
</tr>
<tr>
<td>% Growth</td>
<td>31.19%</td>
</tr>
<tr>
<td>Total Container Revenue (’000 $)</td>
<td>$ 380,204,87</td>
</tr>
<tr>
<td>% Growth</td>
<td>56.17%</td>
</tr>
<tr>
<td>Total Revenues (’000 $)</td>
<td>$ 548,190.62</td>
</tr>
<tr>
<td>% Growth</td>
<td>56.49%</td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td></td>
</tr>
<tr>
<td>Vessel Labor Expenses (’000 $)</td>
<td>$ 259,655.96</td>
</tr>
<tr>
<td>% Growth</td>
<td>30.45%</td>
</tr>
<tr>
<td>% of Revenues</td>
<td>47.51%</td>
</tr>
<tr>
<td>Yard Labor Expenses (’000 $)</td>
<td>$ 72,701.41</td>
</tr>
<tr>
<td>% Growth</td>
<td>143.03%</td>
</tr>
<tr>
<td>% of Revenues</td>
<td>13.31%</td>
</tr>
<tr>
<td>Total Labor Expenses (’000 $)</td>
<td>$ 332,578.37</td>
</tr>
<tr>
<td>% Growth</td>
<td>45.19%</td>
</tr>
<tr>
<td>% of Revenues</td>
<td>60.69%</td>
</tr>
<tr>
<td>M&amp;B Expenses (’000 $)</td>
<td>$ 5,250.00</td>
</tr>
<tr>
<td>% Growth</td>
<td>5.00%</td>
</tr>
<tr>
<td>% of Revenues</td>
<td>0.96%</td>
</tr>
<tr>
<td>SG&amp;A (including Rent) (’000 $)</td>
<td>$ 83,542.88</td>
</tr>
<tr>
<td>% Growth</td>
<td>108.39%</td>
</tr>
<tr>
<td>% of Revenues</td>
<td>15.30%</td>
</tr>
<tr>
<td>Total Expenses (’000 $)</td>
<td>$ 421,411.15</td>
</tr>
<tr>
<td>% Growth</td>
<td>53.66%</td>
</tr>
<tr>
<td>% of Revenues</td>
<td>77.15%</td>
</tr>
<tr>
<td>EBITDA (’000 $)</td>
<td>$ 104,779.47</td>
</tr>
<tr>
<td>% Margin per Move</td>
<td>22.85%</td>
</tr>
<tr>
<td>Capex</td>
<td>0</td>
</tr>
<tr>
<td>Maintenance</td>
<td>0</td>
</tr>
<tr>
<td>Cumulative</td>
<td>0</td>
</tr>
<tr>
<td>Expansion</td>
<td>80,000</td>
</tr>
<tr>
<td>Cumulative</td>
<td>249,705</td>
</tr>
<tr>
<td><strong>Key Operating Metrics</strong></td>
<td></td>
</tr>
<tr>
<td>Vessel Calls</td>
<td>260,761.7949</td>
</tr>
<tr>
<td>Berths</td>
<td>2</td>
</tr>
<tr>
<td>Shc/Cranes</td>
<td>16</td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
</tr>
<tr>
<td>Moves per Customer</td>
<td></td>
</tr>
<tr>
<td>Company 1</td>
<td>833,195.30</td>
</tr>
<tr>
<td>% Growth</td>
<td>211.57%</td>
</tr>
<tr>
<td>Company 2</td>
<td>236,562.09</td>
</tr>
<tr>
<td>% Growth</td>
<td>20.11%</td>
</tr>
<tr>
<td>Company 3</td>
<td>887,103.10</td>
</tr>
<tr>
<td>% Growth</td>
<td>6.89%</td>
</tr>
<tr>
<td>Management Projection</td>
<td>$ 3,294,650.84</td>
</tr>
<tr>
<td>Average Rate Per Move</td>
<td></td>
</tr>
<tr>
<td>Company 1</td>
<td>153.34</td>
</tr>
<tr>
<td>% Growth</td>
<td>33.18%</td>
</tr>
<tr>
<td>Company 2</td>
<td>264.31</td>
</tr>
<tr>
<td>% Growth</td>
<td>42.19%</td>
</tr>
<tr>
<td>Company 3</td>
<td>237.48</td>
</tr>
<tr>
<td>% Growth</td>
<td>40.02%</td>
</tr>
<tr>
<td>Other Revenue per move (optional)</td>
<td></td>
</tr>
<tr>
<td>Company 1</td>
<td>$ 18.46</td>
</tr>
<tr>
<td>% Growth</td>
<td>26.08%</td>
</tr>
<tr>
<td>Company 2</td>
<td>$ 14.18</td>
</tr>
<tr>
<td>% Growth</td>
<td>6.87%</td>
</tr>
<tr>
<td>Company 3</td>
<td>$ 17.18</td>
</tr>
<tr>
<td>% Growth</td>
<td>58.67%</td>
</tr>
<tr>
<td>Management Revenue</td>
<td>$ 157,337,425.36</td>
</tr>
<tr>
<td>Vessel Calls</td>
<td>260,761</td>
</tr>
<tr>
<td>Berths</td>
<td>2</td>
</tr>
<tr>
<td>Vessel Base Wage Per Hour</td>
<td>$ 89.04</td>
</tr>
<tr>
<td>Yard/Gate Base Wage Per Hour</td>
<td>$ 46.29</td>
</tr>
<tr>
<td>Vessel Expense - Labor</td>
<td>$ 258,804,614.40</td>
</tr>
<tr>
<td>Capacity Allotments</td>
<td>0</td>
</tr>
<tr>
<td>Percent of 40’ (or larger) Containers</td>
<td>79.06%</td>
</tr>
<tr>
<td>Gross Acres</td>
<td>382,20</td>
</tr>
<tr>
<td><strong>Model Data (User Input)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Throughout Distribution</strong></td>
<td></td>
</tr>
<tr>
<td>Imports</td>
<td>79.40%</td>
</tr>
<tr>
<td>Exports</td>
<td>19.04%</td>
</tr>
<tr>
<td>Empties</td>
<td>0.36%</td>
</tr>
<tr>
<td>Transhipment</td>
<td>0.28%</td>
</tr>
<tr>
<td>Intermodal Imports</td>
<td>0.81%</td>
</tr>
<tr>
<td>Intermodal Exports</td>
<td>0.15%</td>
</tr>
<tr>
<td>Check</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>Profit</strong></td>
<td>$ 104,779.47</td>
</tr>
</tbody>
</table>

Figure 20 - Simulation Report 2-5
## 7 DISCUSSIONS

In this chapter, the results from theoretical and empirical studies will be discussed. The results of data mining methods are presented and explained in Results Chapter. System creates a report for the user about his/her forecast to analyze and simplify the results and compare different Data Mining methods. This report consists of Correlation and Accuracy percentages of each Data Mining methods. This summary report gives more specific results and chance of comparing all the Data Mining methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjunctive Rule</td>
<td>76.72 - 59.78%</td>
<td>52.72995 - 7.92745%</td>
</tr>
<tr>
<td>IF Rule</td>
<td>80 - 80%</td>
<td>100 - 100%</td>
</tr>
<tr>
<td>ZeroR</td>
<td>0 - 0%</td>
<td>0 - 0%</td>
</tr>
<tr>
<td>Decision Stump</td>
<td>89.24 - 22.88%</td>
<td>52.399985 - 1.1789</td>
</tr>
<tr>
<td>MSR</td>
<td>80 - 80%</td>
<td>100 - 100%</td>
</tr>
<tr>
<td>Additive Regression</td>
<td>100 - 89.58%</td>
<td>98.94395 - 88.9135</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>80 - 80%</td>
<td>100 - 100%</td>
</tr>
</tbody>
</table>

**Figure 21 - Comparison of Data Mining Methods**

The results of several simulations and Moffat & Nichol case show that Additive Regression method is giving more accurate results and relationships between given data. Additive Regression is a MART method which means Multiple Additive Regression Tree. This method includes the use of stochastic gradient boosting which is known for non-deterministic behaviors that user can solve problems under uncertainty. MART can work with all kinds of predictive variables, doesn’t matter quantitative or qualitative. It can handle missing data, predictive variables’ correlation and is robust to the use of irrelevant predictor variables. MART is proved to be an adaptable and suitable tool for building functions without any prior knowledge about relationships between predictors and response variables. [35] All these features make Additive Regression’s results closer to right forecasts.

The results of simulations also showed that Monte Carlo Method is still valid for forecasting purposes even though it is one of the oldest statistical procedures. Monte Carlo analyses are not only conducted by finance professionals but also many other businesses. It proves the concept that every decision has some impact on the overall situation or risk.
8 CONCLUSIONS & FUTURE RESEARCH

In summarizing the thesis, the conclusion is that data mining based technique is an important approach to Port industry. The increasing importance and financial state of marine and port transportation has motivated the research to applying data mining techniques and simulating the Port industry. The development of the tool, IFF, has provided much experience in modeling and easiness in simulating and analyzing a system as complex as Port industry.

The survey of literature indicates that there exists several opportunities for further improvement in Data Mining, not all the areas have been studied by researchers. The use of Data Mining and Monte Carlo provides several interesting models and results to enhance predictions. The results of the Moffatt & Nichol case study suggest that data mining techniques can be useful to obtain hidden relationships between variables and applying several techniques on same data can prevent misunderstanding of models. The advantage of applying multiple data mining techniques on the result is that it forces the analyst to put all the factors that will alter the forecast into an equation but relationships do not mean anything if they are not confident. The tool, IFF, is also assessing the strength of that relationship with root relative square error which is a good indicator of forecasting models’ quality.

In conclusion, IFF is developed for Port industry, but can be extended to other application fields. In the decision making process of firms we can often distinguish several steps. In the first place variables are determined by analysts. Next, a set of old market values and growth direction of markets are being considered, and in such a way that it requires a lot of time and money. Eventually, these values are calculated and presented as forecasts. This tool is developed for the second step. IFF is checking its user’s prediction and presenting some relationships to the user with data mining techniques or simulating forecasts and evaluating risk possibility to its user which saves from time and money. The goal is to provide a methodology and tool for decision makers in port industry so that they can understand and make a better forecast by having an idea about the future.
REFERENCES


Author: Serkan GÜNEŞ
Karlskrona, Sweden
# Table of Contents

IFF MANUAL .................................................................................................................. 41

1. **INTRODUCTION** ................................................................................................. 43
2. **DESCRIPTION** .................................................................................................... 43
3. **DOWNLOAD AND INSTALLATION OF IFF V2.0** .............................................. 43
4. **PAGES** .................................................................................................................. 43
   4.1. **LOGIN** ........................................................................................................... 43
   4.2. **REGISTRATION** ........................................................................................... 44
   4.3. **MAIN PAGE** ................................................................................................. 45
   4.4. **NEW FORECAST** ........................................................................................ 45
   4.5. **APPLYING DATA MINING TECHNIQUES** ..................................................... 46
   4.6. **DATA MINING RESULTS** ........................................................................... 47
   4.7. **DELETE FORECAST** .................................................................................... 48
   4.8. **GENERATE SIMULATION** ........................................................................... 48
   4.9. **LOG OUT** .................................................................................................... 49

# Table of Figures

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFF 1</td>
<td>LOGIN PAGE</td>
<td>44</td>
</tr>
<tr>
<td>IFF 2</td>
<td>REGISTRATION</td>
<td>44</td>
</tr>
<tr>
<td>IFF 3</td>
<td>MAIN PAGE</td>
<td>45</td>
</tr>
<tr>
<td>IFF 4</td>
<td>NEW FORECAST</td>
<td>45</td>
</tr>
<tr>
<td>IFF 5</td>
<td>MAIN PAGE LIST AFTER THE CREATION OF A NEW FORECAST</td>
<td>46</td>
</tr>
<tr>
<td>IFF 6</td>
<td>DATA MINING SELECTION</td>
<td>46</td>
</tr>
<tr>
<td>IFF 7</td>
<td>DATA MINING RESULTS PAGE</td>
<td>47</td>
</tr>
<tr>
<td>IFF 8</td>
<td>GENERATE SIMULATION</td>
<td>48</td>
</tr>
<tr>
<td>IFF 9</td>
<td>MAIN PAGE LIST AFTER THE GENERATING NEW SIMULATION</td>
<td>49</td>
</tr>
<tr>
<td>IFF 10</td>
<td>LOG OUT</td>
<td>49</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

IFF is a Windows® server based data mining and simulation tool that allows you to test, analyze, validate forecasts and generate simulations and analyze them for Marine & Port systems. The foundation of the tool dates back to 2008 [28]. The latest version of IFF is 2.0, which is currently in public beta testing.

Hardware requirements for a server running IFF v2.0 are:

<table>
<thead>
<tr>
<th>Softwares:</th>
<th>XAMPP, WEKA, CutePDF™, Microsoft Excel™</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating System:</td>
<td>Windows</td>
</tr>
<tr>
<td>Download Size:</td>
<td>~15.5 MB</td>
</tr>
<tr>
<td>Installed Program Size:</td>
<td>~80 MB</td>
</tr>
</tbody>
</table>

2. DESCRIPTION

IFF v2.0 is a software application that test, analyzes, validates forecasts and generate simulations as well as generating reports. The tool has a graphical interface where you can upload forecast data. In the tool you will be able to create following:

- *Forecast validation results*, which is generated upon the given forecast variable by user
- *Forecast statistics reports*, statistical user forecast report.
- *Simulation*, which generates a random situation forecast and runs that
- *Simulation validation results*, results based upon data mining techniques
- *Simulation reports*, brief information about the simulation’s facts

The tool is assigned to work on the Microsoft Windows operating system as server; however it is developed as cross platform for client usage.

3. DOWNLOAD AND INSTALLATION OF IFF v2.0

IFF v2.0 is available on CD and can be installed by manually saving and unzipping the IFF v2.0 package on server computer’s hard disk.

4. PAGES

4.1. Login

IFF v2.0 requires password to access the menu. IFF v2.0 is an online tool which means everybody can access from everywhere and create forecasts. User name and password helps system to differentiate users from each other. Another reason is keeping track and results of previous forecasts by users. This will be explained in next sections.
If the user does not have any account in the system, he/she can go to register page by clicking register button on login page. The next thing will be filling the necessary information correctly. If any problem occurs during the registration, system will show what user is doing wrong.
4.3. Main Page

In main page, user can see his/her previous forecasts, simulations and all their results. There are 2 main operations for the user in main page. First one is Creating a new Forecast and applying data mining techniques to the forecast. Second one is Generating a simulation and applying data mining techniques to the simulation.

4.4. New Forecast

First, user needs to give a forecast name in field 1. Next step will be downloading user file to fill. User file can be downloaded by clicking xls file document image in field 2. After the document is filled with user’s numbers, user needs to upload the file into system by clicking “Choose File” button and locating document. After everything is done, user can upload the file and wait for results and statistics report.
At the end of creating forecast result page will be viewed, if there is no problem accepted sign [✓] will be shown, if there is a problem then user will see warning sign [❗]. After these steps, user will be able to see new forecast which was “example1” in this example.

IFF 5 - Main Page List after the creation of a new forecast

- **User File**: User can download the xls file by clicking this icon.
- **Data Mining**: User can click and apply data mining techniques on his/her forecast.
- **Generated Statistics Report**: User can click and download the document easily.
- **Remove**: User can remove his/her forecast and all the results and the report belongs to this forecast.

### 4.5. Applying Data Mining Techniques

In this page, user will be asked for what to apply and evaluate data mining techniques, scenarios or output metrics. Scenarios are user defined cases; Output Metrics are Outputs of Excel.
4.6. Data Mining Results

The structure of this page is same for all Data Mining pages in the system. Doesn’t matter if it is a forecast or simulation. This page provides user, ARFF ( ) file of the forecast and the results of seven different Data Mining methods saved in txt ( ) files. All the results can be downloaded by just clicking on their icons. Also this page shows user a summary report of methods’ results for accuracy and correlation.

 Different types of data mining classifier methods are used to try to predict total profit. Correlation and Accuracy percentage ranges are given for specific methods on this forecast.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination Rules</td>
<td>69.24 - 79.33</td>
<td>952.7922 - 910.3458</td>
</tr>
<tr>
<td>IFM Rule</td>
<td>79.70 - 87.01</td>
<td>978.1105 - 974.5217</td>
</tr>
<tr>
<td>SWAP</td>
<td>90 - 99.99</td>
<td>640.19 - 96.0</td>
</tr>
<tr>
<td>Decision Binning</td>
<td>69.24 - 99.67</td>
<td>972.1795 - 966.5634</td>
</tr>
<tr>
<td>ANF</td>
<td>80.28 - 89.69</td>
<td>989.0245 - 997.8685</td>
</tr>
<tr>
<td>Ant colony Optimization</td>
<td>90.99 - 99.65</td>
<td>997.7415 - 991.4745</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>89.51 - 99.81</td>
<td>990.17625 - 981.70965</td>
</tr>
</tbody>
</table>

IFF 7 - Data Mining Results Page
4.7. Delete Forecast

When user deletes the forecast, forecast won’t be shown in the main list anymore neither its results nor the report. If deletion process is successful, following icon will be appeared on the page for awhile.

4.8. Generate Simulation

The only necessary information is a simulation name for the new generated simulation. After that, tool is applying necessary functions and methods to simulate a forecast.

IFF 8 - Generate Simulation

Generating simulation requires more time than creating a new forecast. So user should be patient and follow the instructions given by tool. IFF will warn user if there is something wrong. When user sees sign on the page, that means simulation is generated without any problem.

After the completion of simulation, it will take its place in the main list, “Simulation2” in our case.
IFF 9 - Main Page List after the generating new simulation

Although the results are different, Data Mining Page is same with IFF 7 - Data Mining Results Page. Simulation Report can be downloaded from PDF sign.

4.9. Log out

If user clicks the “Log Out” button in anywhere of the site, user’s session will be ended immediately. There is no turn back. User needs to login again.

IFF 10 - Log Out