Introducing Collaborative Based Recommendation Systems to the Online Brokerage Domain

Ahmed Eldiwany

A Thesis

In the

John Molson School of Business

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Administration at
Concordia University
Montreal, Quebec, Canada

September 2006

Your file Votre référence
ISBN: 978-0-494-28977-8
Our file Notre référence
ISBN: 978-0-494-28977-8

ABSTRACT

Introducing Collaborative Based Recommendation Systems to the Online Brokerage Domain

Ahmed Eldiwany

Online investment is an industry that utilizes the power and flexibility of the Internet and has enabled investors worldwide to break down the traditional barriers of time and physical location in addition to offering functionalities and opportunities considered unthinkable in the past. The importance of online investment, especially online brokerage, stems from the fact that the various research performed on this topic has demonstrated a rapid increase in the population of online investors, which has not been coupled with the appropriate technological advances providing advice given by professionals in traditional investment and brokerage firms. This lack of advice and recommendations especially in the domain of stock trading has in part motivated this research. The research question being investigated is “What are the effects of introducing collaborative filtering recommendation systems on user satisfaction, user’s impression of usefulness of the system and the intention to use the system in the domain of online brokerage?” In essence, the study examines the user reactions to recommendations provided by technology in the online brokerage sector. Data was collected electronically via the use of two websites designed and coded specifically for the purpose of this research. Results indicate that on average investors using the recommendation system were significantly more satisfied than investors who had no recommendation system. The same investors also reported a higher intention to use the system once it is commercially available.

ACKNOWLEDGEMENTS

I would like to thank Dr. Vahidov for his continued support, guidance and feedback throughout my thesis. Dr. Vahidov was always helpful, open to discussion and encouraging, and without his assistance my thesis wouldn't be where it is today. I thank you dearly for your supervision. I would like to extend my appreciation and thanks to Dr. Abou Zeid and Dr. Kwon for their support and feedback. I am honored to have them all on my committee.

I would also like to thank Dr. Bakr Ibrahim for his continued support and encouragement while working on this thesis.

Finally, I would like to extend my sincere and deep acknowledgement and thanks to my parents, who have provided me with encouragement, motivation and support throughout my work on this thesis and in life. Without their guidance and support, I wouldn't be presenting this thesis today. Thank you for everything you have done for me.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The Internet is undoubtedly one of the most revolutionary technologies invented by man. The introduction of the Internet has allowed businesses to traverse across time and space to reach customers worldwide irrespective of their physical location (Jordan, 2001). Introduced to the public in the 1990s, the Internet today has evolved from an almost entirely static collection of hypertext to a rich amalgam of dynamic services and products that are reachable by millions of users worldwide (Kohrs and Merialdo, 1999). The Internet has changed how businesses communicate, negotiate, operate, market and sell their products to customers. According to a recent study, more than 7,500 terabytes of information have been transmitted online since the introduction of the Internet and more than 513 million users around the globe now have access to the Internet (Miller et al, 2004). From 2004 to 2006, the figure has almost doubled to 1 billion users worldwide, representing roughly 15% of the world's population to date. During the period of 2000-2005, the growth rate of Internet users has been exponentially booming over the 5 year period, averaging an increase of 100% in North America, 400% in Africa, 218% in Asia and 454% in the Middle East (**http://www.internetworldstats.com/stats.htm**)

The Internet and the World Wide Web, however, are not without problems. The Internet may ultimately become a double edged sword. For one thing, the Internet brings diverse opportunities but also carries risks and concerns (Neumann and Weinstein, 1999). One main concern deals with the availability and utilization of investment recommendations in the online medium and specifically for the online investment sector. Despite an abundance of financial information online, online investors in general have very limited access to personalized advice and recommendations, partly due to the fact that recommendations are expected to be customized and tailored according to the investor's portfolio, risk attitude and goals. This poses as a technological challenge to both investors and online brokerage firms as the investor has to make sense out of all the information available on the Internet and must use his or her investment intuition and knowledge to reach the proper investment decisions. In the traditional offline medium however, investors could meet face to face with an investment advisor who manages several portfolios and would customize each portfolio according to the investor's attributes and provide the relevant advice and recommendations to the investor. Nevertheless, not all brokerage firms provide this service, and when they do provide the service it is not for each and every investor, in general it is reserved for investors with a higher investment capital. This is where recommendation systems fit in. Besides the noteworthy benefits of ease of use, comfort, 24/7 availability and lower commission costs provided to investors by online brokerage, recommendation systems offer valuable benefits as well to the firms employing them, including reducing customer search effort, greater customer loyalty, higher sales, an increased advertising revenue, and the benefits of targeted promotion (Mild and Natter, 2002).

Another concern is termed the "Information Overload", which has resulted from the accumulated and nearly unlimited information and data available on the Web. Information overload is a situation that occurs when customers are bombarded with too much information to the extent that instead of facilitating decisions, the increased amount of information leads to confusion and may sway customers away from their objective, which can be harmful especially to E-Commerce. Information overload is one of the most serious issues in today's high tech era, especially in the online domain (Weng et al, 2005). Progressively, customers began turning towards recommendation systems to help them locate the information of interest (Miller et al, 2004). Recommendation systems represent a form of information filtering, an area that is gradually becoming more important nowadays as the amount of information on the Internet is exponentially rising. Recommendation systems refer to the systems "which aim at filtering out the uninteresting items (or predicting the interesting ones) automatically on behalf of the users according to the user's preference" (Cheung, 2003). Recommendation systems have been successfully employed at major websites including eBay.com and Amazon.com, where the recommendation system can provide the users with predicted items of interest based on the items the user selects. The purpose of this study is to introduce collaborative filtering recommendation systems to the realm of online investment, and particularly the domain of online trading in stocks and assessing the feedback obtained from investors regarding the use of this technology to assist them in their investment decisions. The main goals of this research are to create an online prototype of a hybrid collaborative and content based recommendation system that can be applied to provide stock trading advice and recommendations in the online trading environment, as well as permitting the common functionalities of an online stock trading system, while capturing the investor feedback and comments with regards to this approach.

2. Literature Review

2.1. The Internet and the Investor

One significant sector within the Internet realm consists of online investing. Through the various services provided by financial institutions on the Web, users can enjoy the power and comfort of online banking, online bill payments, online procurement and online trading and investment with the major effort being a few mouse clicks. Nowadays, most people tend to trade online through the Internet instead of using full service brokers. Jupiter communications reported that the online brokerage assets in 1998 were estimated at $415 billion, which rapidly compounded more than seven doubles to $3 trillion in 2003 (Fan et al, 2000). Thanks to technology, investors can now sell or buy stocks online, access real time stock market information and statistics, and perform online research using methods that were not even imaginable a decade ago. The diffusion of technology and automation into businesses worldwide has had profound impacts on those markets. However, despite the trend by most firms worldwide to automate their processes, a few markets, including the stock market, are still surprisingly manual despite the vast advancements in technology (Peterffy and Battan, 2004). Additionally, online investment is a major Internet domain which suffers from the information overload problem, due to the existence of a huge quantity of stocks (sometimes over 2000 stocks are available), bonds, futures and statistical financial information regarding each. This appears to be a domain which is in desperate need of information filtering (Tseng, 2004). Online trading firms, such as E-trade and Ameritrade, are among the most successful and vivid financial firms to appear on the Internet within the last decade. The number of online investors was estimated at 12.5 million in over the period from 1995-2000, while it was estimated in 2003 that the number of online investors grew to 42 million, clearly a sharp rise (Dewally, 2003; Barber and Odean, 2001). The potential market for online investment is astronomical; as more and more people worldwide penetrate the Internet population and higher quality and more efficient services are provided. As early as 1998, online trading accounted for 37% of all retail trading volume in equities and options. Nowadays, the percentage is higher and future predictions indicate a further growth of this industry.

The Internet is changing how information is delivered to investors and the ways in which investors can act on that information. However, due to information overload, investors may not be able to make complete sense of that information or reach a decisive conclusion on which stocks, bonds or futures to purchase. Many of today's investors are new to the market and act on instinct, intuition or "word of mouth" advice. This situation can lead to a sort of "dark side" in the online investment domain. Since investors can now purchase or sell any equity with a few mouse clicks, investors gain an exaggerated sense of control over the outcome of their trades. Moreover, the nearly unlimited information supply via the Internet can lead to overconfidence in trading behavior. Rapid selling and procurement of equities, coupled with faster feedback, direct the investor's attention on recent performance (representing short term vision), which can have negative consequences on the investor's returns and profit margins over prolonged periods of time. (Barber and Odean, 2001). Online brokerage replaces people and telephones with computers and code. Online brokerage firms have numerous advantages over traditional brokerage firms, including much lower set up costs for establishing the online brokerage firm, cheaper and faster interaction with the users, lower marginal costs, 24/7 service availability, the ability of the firm to work from a centralized location while allowing the investor global reach (via the use of the Internet) and the need for less skilled labor and human employees. For instance, Merrill Lynch, a full service traditional brokerage firm employs 20,200 financial consultants and other professionals to manage 124,000 commission generating transactions daily, while E-trade, an online brokerage firm, employs 2,800 professionals and manages 283,000 commissions generating transactions per day on average, a number that is almost double that of Merrill Lynch, and achieves this with merely 10% of the manpower. For investors, online brokerage can also provide various advantages including the availability of free advice and information on the Internet, lower trade commissions, faster trade execution, and the ability to utilize the power of recommendation services when applied to the online investment sector, which is the main aim of this study. However, in this environment, investors using the online brokerage services find themselves depending on their own skills and knowledge and without the guidance and assistance of a personalized broker with which the investor can interact with, emphasizing the need of a "replacement" service which can be demonstrated in recommending stocks (or other equities and assets) to the investor based on certain attributes that characterize the investor and based on the current market conditions and equity prices (Barber and Odean, 2001). Clearly, this domain represents one among many online domains which is in critical need for the services of recommendation systems

Personalization is the wave of the future in both the online and offline environments. Recent studies have demonstrated that customer's demands are increasing and evolving. A single online configuration for all users will likely fail to achieve wide support and satisfaction since users have different backgrounds and different goals they wish to achieve (Kohrs and Merialdo, 1999). Gradually, more and more websites are attempting to build closer relationships with its customers by "adapting" to their needs and aim at providing a more personalized experience. Recommendation systems fit into this picture by providing recommendations and presenting information based on the distinct interests of each individual user, which requires the use of powerful filtering technology to achieve. Utilizing this technology, both firms and individuals cut down the time required for users to locate information, services or products and make decisions, such as purchasing decisions, regarding these products. One of the most successful technologies available for recommendation systems is known as collaborative filtering. Collaborative filtering essentially uses technology and certain mathematical search techniques to match the similarities and dissimilarities among different customer tastes and preferences to offer recommended products or items to the target customer (Srikumar, 2004). This technology has been improved to a point where there exists a wide array of methods and algorithms for providing recommendations based on the necessary goals of the system (Herlocker et al, 2004).

2.2. Electronic Brokers: Evaluation and Selection

Throughout history, banks have played the major role of money keepers and financial management for most people in the world. The Royal bank of Canada, for instance, is the nation's largest financial institution to date (Choo and Johnston, 2004). However, this scenario might be altered sooner than most people think. New financial institutions relying on a technological base, such as electronic brokers, are threatening to alter the current financial market status quo. During 1999 and 2000, 16% of all stocks traded in North America were performed by electronic brokers with the numbers on the rise. Currently, this percentage is estimated at 29%, while simultaneously the personal online trading account is rapidly becoming the main personal financial gateway for many users, replacing the traditional bank. The market share online brokers gained in the daily online trades rose from zero in 1995 to nearly 73% in 2003, with the numbers still climbing (Levinsohn, 1999). Switching to online trading provides users with a lower average price per trade, and this benefit is the most relevant factor driving users to switch to online trading. For instance, in 1995, prior to the introduction of online trading, the average price per trade was $73. In 2003, after the diffusion of online trading to users from a variety of backgrounds, the average price per trade dropped to from $73 to $16 (roughly an 80% drop in cost) and is still in decline. Moreover, a larger number of small and medium sized investors have entered the stock market and seek lower cost services provided by discount brokers rather than the expensive and customized full service brokerage services provided by firms like Merrill Lynch (Levinsohn, 1999; Saatcioglu et al, 2001). With technology rapidly becoming more affordable, and with the rapid proliferation of the Internet into the lives of millions of users worldwide, the potential market for online trading is astounding. Despite the existence of dozens of online brokerage firms in North America, Asia and Europe, many regions in the world including the Arab world and Africa still provide very limited online brokerage services. These are markets that have a real potential for near future growth due to the fast pace of Internet proliferation into those regions (about 400% increase in Internet usage in Africa and 454% in the Arab world over the past 5 years). Despite its numerous advantages, online trading has its own drawbacks. One core problem that faces investors who switch to online trading is the lack of personalized recommendations and advice, which can be a very influential factor in determining the investor's portfolio management decisions especially for the less experiences investors. Despite obtaining technological benefits such as being more in control and purchasing stocks for lower prices and possibly at a faster execution time, most investors still lack the resources to make high quality investment decisions on their own (Saatcioglu et al, 2001). This is where well designed recommendation systems for online investment can align themselves perfectly, providing low cost, high quality and solid recommendations for purchasing decisions which incorporates the stock history, the risk profile of the investor, the goals and aims of the investor, what similar investors purchased or a variety of other determinants.

For a private electronic stock trading system to function properly, it must be interfaced into the major Electronic Communication Networks (ECN) such as the NASDAQ (Maxemchuk and Shur, 2001). This is to allow automated order execution based on the commands provided by the investor, who interacts with the system and places the orders online via a web browser such as Internet Explorer (Konana et al, 2000). Since most firms have the same access to technology, stock market research and statistics concerning consumer behavior, it would be insightful to understand how potential investors evaluate electronic brokers and what are the steps investors implement to ultimately decide which online brokerage service to use. Commenting on this issue, Konana et al (2000) identified that ease of use, quality of the research, commission rates, customer service and reporting abilities were major factors impacting investor's evaluation of an electronic broker, with less investor attention to be directed towards less obvious factors such as the ability to obtain the best possible stock prices and the timeliness of the transaction. These factors are summarized and presented in **Table 1** below. It is critical to note that the mentioned factors are based on the investor's "perceived" efficiency of the electronic broker not the "actual" efficiency. Since generically decisions are based on perceptions and subjective views, it is worthy to note this distinction. According to Konana et al (2000), "perceived" efficiency in this sense is generally determined by information that can be verified, such as commission rates, research material etc. In their model, perceived efficiency is moderated by the investor's risk attitudes and the degree of trust the investor assigns to electronic brokerage firms.

1. Ease of use of the system
2. Quality of the research information provided
3. Commission rates and commission structure
4. Customer service
5. Reporting abilities provided by the firm
6. Best possible stock purchase price
7. Timeliness of executing the transaction

Table 1: Factors impacting investor's evaluation of E-brokers

After successful evaluation of the electronic broker, investors must then select the broker that best meets their needs. Based on research in consumer behavior and economics, Konana et al (2000) propose a model for electronic broker selection composed of 9 steps, which is provided in a shortened form in **Table 2**. The first selection step involves the investor assigned "weights" to the attributes of the electronic broker that the investor considers important (such as the ease of use, the commission fees etc.). The second step involves the investor performing preliminary research for information regarding the electronic brokers, how they operate and what kind of information they provide in an attempt to narrow down the search to one or two electronic brokers. The third step consists of investors studying the information from the electronic brokers, including the terms placed by the brokers, their research quality, online transaction fees and online advice. The investor gathers this information from multiple sources that can include surfing the websites of e-brokers, obtaining recommendations from friends and family, or other sources. The fourth step involves the investor "calculating" the value of the verifiable attributes of the electronic broker, such as the trading cost, in order to arrive at a final decision on which broker to select. The fifth step consists of the investor forming an expectation of values for the unverifiable attributes, those attributes which are not obvious at first glance. The level of trust the investor extends to the electronic broker is an integral part of this step. The sixth step involves the investor adjusting expectations about the unverifiable attributes to accommodate the investor's risk attitudes. This is the step that involves uncertainty as unverifiable attributes (or the attributes that are non-measurable and hence difficult to assess) are more fluid and flexible than verifiable ones, and more difficult to estimate. The seventh step consists of determining the overall value for each electronic broker, and ultimately selecting the one that provides the investor with the maximum value while meeting all the investor's needs. The eighth step of the selection process regards modifying the knowledge structure with information gathered during usage. In other words, after initial use of the service, investors begin to better understand the value and quality of the unverifiable attributes of the electronic broker and therefore can reassess their judgment regarding that broker. Finally, the ninth and last step of the process involves searching for alternative electronic brokers if the overall impression is unsatisfactory.

1. The investor assigned "weights" to the attributes of the electronic broker that the investor considers important
2. The investor performs preliminary research for information regarding the electronic brokers, how they operate and what kind of information they provide
3. The investor studies and examines the information from the electronic brokers, including the terms placed by the brokers, their research quality, online transaction fees and online advice
4. The investor "calculates" the value of the verifiable attributes of the electronic broker, such as the trading cost, in order to arrive at a final decision on which broker to select.
5. The investor forms an expectation of values for the unverifiable attributes, those attributes which are not obvious at first glance.
6. The investor adjusts the expectations about the unverifiable attributes to accommodate the investor's risk attitudes. This is the step that involves uncertainty as unverifiable attributes (or the attributes that are non-measurable and hence difficult to assess) are more fluid and flexible than verifiable ones, and more difficult to estimate.
7. The investor determines the overall value for each electronic broker, and ultimately selecting the one that provides the investor with the maximum value while meeting all the investor's needs
8. The investor modifies the knowledge structure with information gathered during actual usage
9. The investor searches for alternative electronic brokers if the overall impression after usage is unsatisfactory.

Table 2: Steps for E-broker Selection

2.3. The Online Trading Process

In their research on the Internet and the future of financial markets, Fan et al (2000) propose an extremely informative framework depicting the key players in the online financial trading process and the directions of order flows. The model is presented in **Figure 1**. As can be seen in the figure, the online trading process is not a single path model. The various market centers such as the NYSE, ECNs (Electronic Communication Networks) and market makers all compete with each other for order flows from investors. Each market utilizes a different trading mechanism to match the buy and sell orders. This unique mixture enables diversity in market outcomes in terms of the fees charged on the transactions and the timeliness of those transactions. Brokerage firms are a key player in the process. Brokers in broad terms make order routing decisions, selecting which path to follow for the order flow. This decision usually is not in the hands of the investor, who is generally unaware of the details behind the transactions. The main players in the market able to absorb the investor's order and execute the order include the exchanges, the market makers, the ECN and internationalization:

1 Exchanges: There are a number of well known, traditional stock exchange markets around the globe where a large portion of the trading process is executed. The most famous exchange is probably the NYSE (New York Stock Exchange).

2 Market Makers: Market makers can be of two main types (in the United States). NASDAQ is the first type of market makers, which are significant. The NASDAQ market makers are dealers who are willing and prepared to purchase or sell stocks

traded in the NASDAQ market. The second type is the third market makers, who are dealers that buy and sell stocks listed on an exchange at publicly quoted prices

3 Electronic Communication Networks (ECN): ECN represent a modern introduction of technology to the stock market. ECN are defined to be "Electronic auction systems that can execute trades without human intervention" (Fan et al, 2000 p.85). ECNs already account for nearly 30% of total share volume traded on the NASDAQ market. In broader terms, ECN account for 3% of the total share trading volume of stocks that are exchange listed. ECN provide the benefits of being cheaper and faster in execution than regular, organized exchanges, which in turn can be viewed as a potential competition for those markets

4 Internationalization: Certain market makers can be affiliates or partners with a brokerage firm. Brokerage firms can directly route the order to an affiliate market maker anywhere in the world (Fan et al, 2000).

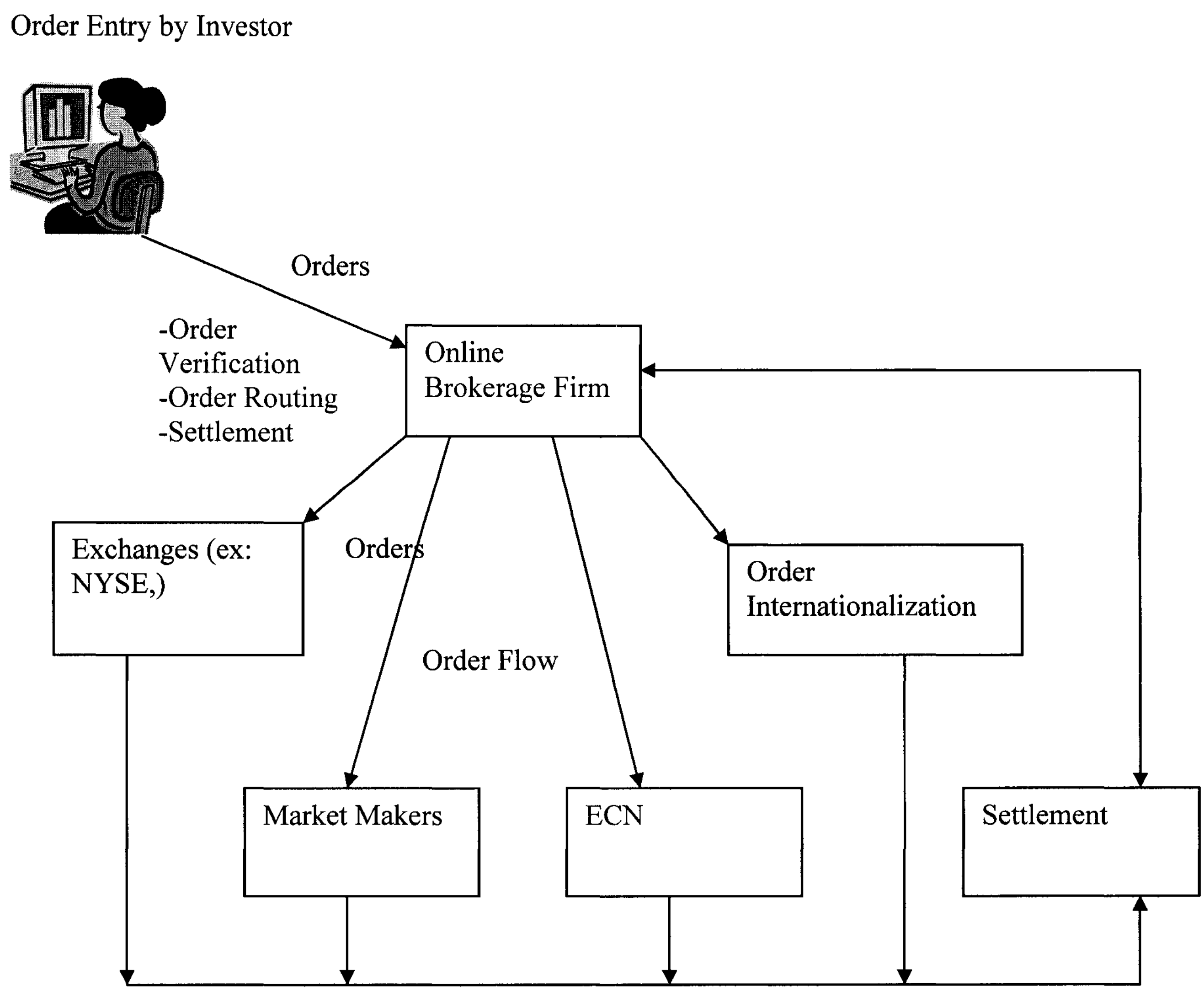


Figure 1 : Model by Fan et al (2000) for the Online Financial Trading Process

3. Recommendation Systems and Collaborative Filtering

3.1. Overview

As previously mentioned, recommendation systems are defined to be systems that "use the opinions of a community of users to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices" (Herlocker et al, 2004). Recommendation systems offer valuable benefits to firms employing them, including reducing customer search effort, greater customer loyalty, higher sales, more advertising revenue, and the benefits of targeted promotion (Mild and Natter, 2002). **Table 3** on the following page summarizes the benefits of recommendation systems to firms as well as some drawbacks.

There appears to be a wide consistency among the literature regarding the definition of recommendation systems, and the goals that these systems aim at achieving. For instance, Konstan (2004) defines recommendation systems as "systems that use the opinions of members of a community to help individuals in that community identify the information or products more likely to be interesting to them or relevant to their needs", a definition very similar to that of Herlocker et al (2004). The most common example of recommendation systems can be demonstrated in content filtering approaches, which aim at filtering information using "keywords" to match the relevant documents. Content based recommendation systems extract an item's characteristics and compares it with the user's interest profiles for predicting the user's preference over the items (Cheung, 2003).

Table 3: Definition, Benefits and Drawbacks of Recommendation Systems

Recommendation Systems: Systems which aim at filtering out the uninterested items (or predicting the interesting ones) automatically on behalf of the users according to the user's preference" (Cheung, 2003)

Benefits:
• Increased Personalization to users • Mimics "word of mouth" recommendations to the online environment • Provides assistance to users of E-commerce applications who previously had to rely on their own knowledge for all activities online • Reduces time and cognitive effort users need to locate items of interest • Helps users predict possible items of interest users were not aware of • Higher customer satisfaction • Increased customer loyalty • Greater customer retention rates and re-purchase rates • Convenient and fun to use • Offers firms the benefit of targeted promotion
Drawbacks:
• Different types of recommendation systems yield varying results according to how it is applied (i.e. non standard results) • Results may be meaningless to the user (recommendations may be inaccurate) • Requires explicit customer ratings to function accurately • Accuracy suffers from data sparsity (lack of data) • Not applicable to all E-commerce applications • Many recommendation systems do not provide explanations as to why a recommendation is being made and how the recommendation was reached by the system • Requires users to trust the recommendation system and the intentions behind the recommendations

A good example is www.Google.com, which can locate documents of interest to the user via the use of content based "keyword" filtering and returning the relevant documents of potential interest to the user. Another approach, which happens to also be one of the most successful technologies employed in order to achieve this information filtering is termed collaborative filtering, an approach introduced in 1992 by Goldberg et al and it is the approach that will be examined in more detail throughout this study (Ansari et al, 2000; Goldberg et al, 1992). *Collaborative filtering* is defined to be "an approach to make recommendations to users by identifying correlations among ratings given to items by a group of users" (Lin et al, 2004). Collaborative filtering was developed to address two essential challenges that could not be solved by available content based information filtering techniques. First of all, collaborative filtering systems were designed to address the problem of overwhelming quantities of documents that were on topic. Secondly, they address the concern of filtering non-textual documents based on human judgment and taste, and not merely based on keywords (Konstan, 2004). In the simplest form, collaborative filtering "predicts a person's preference as a weighted sum of other people's preferences, where the weights are proportional to correlations over a common set of items evaluated by two people" (Ansari et al, 2000). Essentially, collaborative filtering techniques match the similarities and dissimilarities among different customer tastes and preferences to offer recommendations to the target customer (Srikumar, 2004). One of the main purposes behind the creation of collaborative filtering is the need for firms to provide some sort of recommendation and help to its online customers. In traditional brick and mortar stores, employees provide customers with information regarding services, products, concerns, questions they have about products or explain promotion activities. In the online environment, the situation is more complex. The lack of an "agent" to interact with the customer can have a negative impact on the business, hence novel methods such as intelligent agent techniques are used to support such activities as information collection, filtering and recommendation (Lin et al, 2004). Specifically, collaborative filtering provides three fundamental benefits to firms and users including the ability to filter any type of content (such as text, art, music, equities and movies), the ability to filter based on complex and hard to represent concepts such as taste and quality, and the ability to make unexpected recommendations to users (Herlocker et al, 2000). Collaborative filtering is a personalization tool used by firms to provide a form of customized service to customers. Personalization, which is defined to be the use of information about a particular user to provide tailored user experiences, is the new trend in the online and offline environments (Alpert et al, 2003). Firms that offer user centered, easier to use, personalized web pages and services will have a competitive advantage over firms that do not offer these services. In other words, recommendation systems in the online environment have become more of a necessity than a luxury (Srikumar, 2004). Customers are gradually being bombarded with an overwhelming magnitude of information that is quickly becoming an information overload problem in even the simplest purchases. Studies have demonstrated that customers have become more time starved as well, which emphasizes the need for improved recommendation systems, especially in the online investment domain as it lacks such a system to date.

3.2. Collaborative Filtering Approaches and how they operate

A summary of the types of collaborative filtering approaches, along with their respective benefits and drawbacks, is presented in **Table 4**. Collaborative filtering predicts user preferences for items in a "word of mouth" manner. That is, the preferences of each unique user are determined by considering the opinions (usually in the form of explicit preference rating) of other "like minded" users (Cheung, 2003; Jin and Si, 2004). Collaborative filtering operates similarly to an offline social network, where a user or customer asks friends, family and colleagues to provide recommendations regarding a product or service they have experience with and whom the user trusts to have "similar" opinions from previous interaction experience. Generically, there are two main approaches to collaborative filtering that are addressed in this study, known as memory or model-based methods. Other collaborative filtering approaches include Maximum Entropy and Association Analysis, but are not discussed in detail in this study, and a detailed explanation of those types can be found in Jin et al (2005); Zitnick and Kanade (2004). In general, collaborative filtering requires a database with existing ratings for items in order to provide accurate recommendations. In order to obtain the ratings initially, Ansari et al (2000) have identified 5 main sources of recommendations for items:

1. A user's expressed choice or explicit rating among alternative products
2. Preference for product attributes
3. Other people's preferences or choices
4. Expert judgments
5. Individual characteristics that may predict preferences

Table 4: Summary of CF approaches, benefits and drawbacks

Collaborative Filtering(CF): an approach to make recommendations to users by identifying correlations among ratings given to items by a group of users" (Lin et al, 2004). Essentially, collaborative filtering techniques match the similarities and dissimilarities among different customer tastes and preferences to offer recommendations to the target customer (Srikumar, 2004).

Model-Based CF: The algorithm groups different users in the database into a small number of classes based on their rating pattern. For the prediction to occur, the algorithm first categorizes the active user into at least one of the pre-defined classes and uses the ratings of those predicted classes on the target item as the prediction

Benefits:

- Calculations performed on a cluster instead of the entire database saving computational time and leading to faster calculations
- Faster prediction generation
- More accurate than memory-based methods in data sparsity situations

Drawbacks:

- Generally less accurate than memory-based methods
- Poor coverage compared to memory-based approaches

Memory-Based CF:

These methods first compute the similarities between users by comparing directly their preference rating (which is usually on a Likert type scale over a range of 1-N). The preferences of a user for an unrated item is then predicted by summing up the contributions of other users for the same item, and weighted based on a user similarity measure (Cheung, 2003).

Benefits:

- Generally more accurate than model-based collaborative filtering methods
- Generally better coverage than model-based methods

Drawbacks:

- Performs complex computations on the entire database
- Suffers from data sparsity potentially leading to poor recommendation accuracy
- Difficulties in data scalability as the growth in computations is exponential as more data is introduced
- Possibly high latency (computational) time

3.3. Memory-Based Collaborative Filtering Methods:

These methods first compute the similarities between users by comparing directly their preference rating (which is usually on a Likert type scale over a range of 1-N). The preferences of a user for an unrated item is then predicted by summing up the contributions of other users for the same item, and weighted based on a user similarity measure (Cheung, 2003). Memory-based approaches works by identifying the similarity between two users by comparing their rating over a set of items. These algorithms undergo complex computations on the entire database to identify the K top most similar users to the active user in terms of patterns and then combines those ratings together. The core rationale behind memory-based algorithms is to compute the active users predicted rating of an item as a weighted average of ratings by other K similar users or K nearest neighbors (Xue et al, 2005). The most common methods are neighborhood based approaches (Vucetic and Obradovic, 2005). Some of the most popular examples of memory-based algorithms include Pearson's Correlation Coefficient and Vector Similarity approaches. In its simplest form, memory-based methods perform the following 3 steps to produce a recommendation:

1. The user enters the profile of ratings
2. The collaborative filtering systems locates users with similar profiles (neighbors)
3. The ratings of neighbors are combined (using one of numerous algorithms in the literature) to form a recommendation.

These steps are decomposed into smaller sub-steps during implementation. However, the users are generally not aware of such details. Step 2 is a critical step for recommendation systems, and this is where various algorithms have been proposed to increase accuracy and decrease running time necessary for locating users with similar preferences to the target user. Step 3 is where the actual rating prediction is aggregated, and it is this during this step that several researchers have proposed introducing explanations to tackle weak predictions and increase the user's trust in the system (Herlocker et al, 2000).

3.4. Evaluating Memory-Based Collaborative Filtering Approaches:

Memory-based approaches have been found to suffer from two fundamental problems, namely data sparsity and difficulty in data scalability. Data sparsity refers to the situation where most users only rate a small percentage of items that they come across; hence a very sparse user item matrix is available for providing recommendations, leading to poor recommendation accuracy. Difficulties in data scalability refer to the exponential growth in the required computations as the recommendation systems needs to compare user profiles to identify similarities before making a recommendation, which requires heavy computational demands that grows with both the number of users and items (Lin et al, 2004). These two problems can often lead to poor recommendation accuracy, and have inspired researchers to find alternative approaches to avoid the pitfalls of these problems. On the other hand, memory-based approaches have been found in studies to be more accurate than model-based approaches. However, their high latency (computational time) in giving predictions to the interested user can pose as a serious drawback with large systems consisting of vast amounts of requests that need to be processed in real time (Vucetic and Obradovic, 2005)

3.5.Model-Based Collaborative Filtering:

The second approach to collaborative filtering, known as the model-based method, approaches the recommendation problem in a different manner. The algorithm groups different users in the database into a small number of classes based on their rating pattern. For the prediction to occur, the algorithm first categorizes the active user into at least one of the pre-defined classes and uses the ratings of the nearest class on the target item as the prediction. Notable examples in this category include Aspect Models, Bayesian Networks, and Clustering Approaches (Xue et al, 2005). The clustering approaches, for instance, works in the following manner. Firstly, groups of users who seem to have similar preferences are identified and clusters are created for each group of similar users. Secondly, predictions for a user can be made by averaging the ratings of the other users in that cluster only (as opposed to scanning the entire database in memory-based methods) and assigning it to the user.

3.6.Evaluating Model-Based Collaborative Filtering Methods:

The true power of model-based approaches can be observed in the fact that they alleviate the serious drawbacks of memory-based approaches, specifically the data sparsity and data scalability problems (Xue et al, 2005). Model-based approaches in general have been demonstrated to be less accurate than memory-based methods, although they have faster computational time as the entire database is not searched in real time for the K top most similar users as in the memory-based methods (Vucetic and Obradovic, 2005). However, model-based approaches have also been found to be more accurate in data sparsity situations than memory-based approaches, which suffer heavily under such conditions (Cheung, 2003). Some studies have also shown that model-based methods have a poor coverage when compared to memory-based methods. However, the same reasons that characterize model-based methods as having poor coverage also justify its lower latency time and faster prediction generation. It appears that there is no "best method" for constructing collaborative filtering systems. Each method or approach has to be evaluated according to its advantages, drawbacks and the required task to be performed. Some algorithms are developed for certain tasks and should be used for those tasks only. Assessing and selecting the proper collaborative filtering approach is a design decision that must be made for each system that is being built, depending on the systems' requirements.

3.7. Robustness of Collaborative Filtering Systems

One important issue mentioned in the collaborative filtering system research regards the robustness of these systems. Two fundamental concerns fall under the robustness category, namely the errors that occur during prediction estimation and the rating variance among the different users of the system. Generically speaking, Herlocker et al (2000) identify the following two main categories of errors in collaborative filtering systems:

1. Model and process errors: These errors occur when the system uses a process to compute recommendations that does not match the user's requirements
2. Data errors: Are errors that occur due to the sparsity and inadequacy of data used to compute the recommendation. Data errors include
    a. Not enough data (data sparsity)

b. Poor data

c. Variance in the data

Regarding the rating variance problem, Jin and Si (2004) have identified two main types of rating variances among users:

1. Shift of Average Ratings: This rating variance problem occurs when some users are more "tolerant" than other users when assigning ratings. This implies that on average, the "tolerant" users will tend to give higher ratings than others. The consequence of such actions can be observed in a higher average rating for those users. Jin and Si (2004) propose addressing this issue by subtracting ratings of each user from the average ratings of the same user (a process also known as normalization).
2. Different rating scales: This rating variance problem, on the other hand, occurs when some users are more "conservative" than other users, and tend to provide ratings within a narrow range of rating categories, whereas "liberal" users tend to assign ratings within a much wider range of ratings. To address this concern, it was suggested by Jin and Si (2004) to divide the ratings of each user by the variance in the same user's ratings (another form of normalization).

It should be noted that there are several other limitations to collaborative filtering, irrespective of the approach or methodology that is selected. In the discussion regarding the constraints on collaborative filtering systems, Ansari et al (2000) identify the following 4 additional factors in besides those already mentioned in this paper:

1. Collaborative filtering can only be used when preference or rating data for an item already exists in the database
2. Collaborative filtering tend to use ad-hoc algorithms, and thus do not account for uncertainty
3. Collaborative filtering systems do not explicitly incorporate attribute information
4. Collaborative filtering systems are co-relational; they provide little explanation as to why the recommendation is being made, a factor that is essential for trust building.

3.8. Security of Collaborative Filtering Systems

Systems that automatically adjust themselves to input from users are vulnerable to attack from those same users. Recommendation systems in particular are a common, easy target for attacks for a multitude of reasons, including political, financial or other motivation factors that encourage the promotion or demotion of recommended items (O'Donovan and Smyth, 2006). Research from various sources has demonstrated that incorporating trust reputation and explanation feedback can have a positive impact on the robustness, accuracy and user acceptance of recommendation systems (O'Donovan and Smyth, 2006; Miller et al, 2004; Pu and Chen 2006; Dellarocas 2006). Only recently has the subject of collaborative filtering system security been brought to the spotlight. Most of the literature on collaborative filtering discusses performance of the systems rather than security concerns (O'Mahony et al 2006). However, it appears the issue of security will be a critical concern for both system designers and system users. Since a user's recommendation is usually calculated based on what other "unknown" individuals have rated, it is a difficult and tiresome task to distinguish between actual, "real" ratings and those ratings produced maliciously. Additionally, in order to perform an extremely successful attack, attackers only need to know a very small amount of information about the system (Lam and Riedl, 2004). This makes the task of protecting the collaborative filtering system even more difficult.

Noise in itself is a result of "imperfect user behavior and various collection processes that are employed" (O'Mahony et al, 2006). Two main categories of concerns are considered for recommendation systems, namely natural noise and malicious noise (intentional attacks). Natural noise is a class of noise which relates to the methods by which collaborative filtering systems obtain user preference and ratings. Since humans are subject to error, natural noise should be expected to occur. Certain algorithms and methods exist in the literature for removing natural noise, and hence it is not a major concern for collaborative filtering systems. On the other hand, malicious noise, also known as attacks, is a much more serious consideration. This type of noise represents biased noise deliberately placed by an attacker in order to impact the ratings and recommendations of certain products or services. Attackers have a wide array of methods by which to attack a collaborative filtering system. Attackers can form several online identities through multiple registrations with the system. Each of these separate identities is known as an attack profile. In broad terms, there are two fundamental types of attacks; product push attacks and product nuke attacks. The different types of malicious attacks are summarized in **Table 5**. O'Mahony et al (2006) have identified the following 4 attack strategies for product push attacks:

1. Random attack: In this type of attack, attack profiles are created by choosing (N-1) items at random from the entire product space. The item being "pushed" is incorporated in the profiles.
2. Popular Attack: In this attack type, the rationale used by the attacker is that the most popular items available in the product space are good candidates for an attack. Such items are easily identifiable and provide some benefits to attackers. The benefits include that such popular profiles have a high probability of being located in the neighborhood of many users, hence reducing the cost of an attack in terms of size and number of attack profiles required.
3. Probe Attack: This type of attack utilizes the output of the recommendation system in order to determine which attack profile to select. This type of attack is difficult to uncover as by rating a small number of initial "seed" items, an attacker can gradually build up attack profiles that will closely match the distribution of genuine users of the system.
4. AverageBot Attack: This attack type involves the creation of an attack profile consisting of all the items in a system. This is an aggressive attack aiming to influence the entire system. This attack can be critical, and for the stability of the business employing the collaborative filtering technology proper precaution measures must be taken to avoid such an attack.

In their discussion on the various security and attack threats that are potentially harmful for a recommendation system, O'Donovan and Smyth (2006) and Burke et al (2005) discuss other attack types, with some being similar to those O'Mahony et al (2006) discussed in their research on security of recommendation systems. These additional four attack types include:

5. Shilling Attack: is defined to be an attack by a malicious user or set of users that are trying to alter the behavior of the system to suit their own needs
6. Sampling Attack: Is an attack in which the attack profiles are sampled from real user profiles in the database, making the attacking profile difficult to determine.
7. Bandwagon Attack: This type of attack inserts the most popular items into the attacking profile (very similar if not identical to the Popular attack discussed above)
8. Favorite item Attack: This type of attack is directed not towards the entire system but rather towards a particular user. It is assumed the attacker has a general idea of the user's taste preferences.

Table 5: Types of malicious attacks on RS

1.	Random attack: In this type of attack, attack profiles are created by choosing (N-1) items at random from the entire product space. The item being "pushed" is incorporated in the profiles.
2.	Popular Attack: In this attack type, the rationale used by the attacker is that the most popular items available in the product space are good candidates for an attack. Such items are easily identifiable and provide some benefits to attackers. The benefits include that such popular profiles have a high probability of being located in the neighborhood of many users, hence reducing the cost of an attack in terms of size and number of attack profiles required.
3.	Probe Attack: This type of attack utilizes the output of the recommendation system in order to determine which attack profile to select. This type of attack is difficult to uncover as by rating a small number of initial "seed" items, an attacker can gradually build up attack profiles that will closely match the distribution of genuine users of the system.
4.	AverageBot Attack: This attack type involves the creation of an attack profile consisting of all the items in a system. This is an aggressive attack aiming to influence the entire system. This attack type is crucial for the stability of the business employing the collaborative filtering technology, and the proper precaution measures must be taken.
5.	Shilling Attack: is defined to be an attack by a malicious user or set of users that are trying to alter the behavior of the system to suit their own needs
6.	Sampling Attack: Is an attack in which the attack profiles are sampled from real user profiles in the database, making the attacking profile difficult to determine.
7.	Bandwagon Attack: This type of attack inserts the most popular items into the attacking profile (very similar if not identical to the Popular attack discussed above)
8.	Favorite item Attack: This type of attack is directed not towards the entire system but rather towards a particular user. It is assumed the attacker has a general idea of the user's taste preferences.

In order to envision and grasp the security threats from potential attacking profiles, it would be useful to provide an example of how an attacker would go about manipulating the system. A very good and basic example is provided in the study by O'Donovan and Smyth (2006). Let us assume that there are 10 attacking profiles in the current recommendation system, which are called p1....p10. During the trust building process, user p1's ratings are computed by allowing user p1 to generate ratings (predictions) for the other users in the system, which include the remaining p2...p10 profiles. Assuming that attacking profiles more or less follow similar rating patterns, profiles p2...p10 will reinforce the ratings provided by user p1 resulting in a higher "trust" value for that attacking profile. This could have a consequence that in the final recommendation process, it is possible that the opinions of the attackers will carry more weight, and thus impact, on the system rather than the "real" profiles. This undeniably is a situation of great concern for recommendation system designers and users, as it shakes the trust users have in the system and poses a serious security challenge to the system designers and administrators.

Before continuing the discussion on attack profiles it appears important to explain the definition of some security terminology to be covered. A *prediction shift* is the difference in the rating of an attacked item before and after the attack has been successfully carried out. In other words, the prediction shift is actual change (or shift) that occurs to the recommendation of an item as a result of an attack. *Hit Ratio* is defined to be the percentage of time a targeted item gets into the top N recommendations by an attack (O'Donovan and Smyth, 2006). Luckily, there are methods by which the administrators of recommendation systems can limit the influence of attacking profiles. In the example scenario provided above, the solution to the attacking profile's manipulation of the ratings is to initially modify the manner by which the profiles are selected to produce recommendations and not opening the door to allow recommendations by all users. By using certain deterministic metrics such as the length of time a user has been with the system, the reliability of that user profile's ratings, whether or not the user is registered with the system and other similar measures can provide a reduced prediction shift of an attack by up to 75% (O'Donovan and Smyth, 2006). Hence, it is preferred to accept recommendations into the system only from users whom the system "trusts".

Once an attacking profile is ready to bombard a recommendation system with false ratings, there are four main scenarios identified by Dellarocas (2006) that the attacker can utilize to influence the system ratings, depending on the desired outcome:

1. Unfairly high ratings (ballot stuffing): This approach is used when a seller, for instance, plots with a group of buyers and provides them with unreasonably high ratings in order to inflate the reputation and encourage more orders from those particular buyers and possibly at higher prices as well.
2. Unfairly low ratings (ballot mouthing): Sellers and buyers can conspire in order to "bad mouth" the competition or other sellers/buyers in order to drive them out of the market. This is equivalent to the modern day use of "scandals" and other means to

discourage activity with certain user and is a means to provide a biased negative rating.

3. Negative Discrimination: In this situation, sellers for instance can provide excellent services to all users except an unlucky few whom the seller does not like (such as the competition).
4. Positive Discrimination: In this scenario, sellers or buyers can provide poor service to all users they interact with except a lucky few, who receive excellent service.

In order to avoid the side-effects of such manipulation in the ratings, recommendation system designers and administrators have in essence two approaches to deal with this difficult situation. The first approach, called *avoidance mechanisms*, can pro-actively seek to prevent malicious ratings from entering the system initially. The second approach, known as *recovery mechanisms*, can be utilized to detect occurrences of malicious user behavior and attempt to minimize the harmful impact on the system. Dellarocas (2006) proposes Avoidance mechanisms consisting of the following 3 approaches to prevent user malicious manipulation to system ratings:

1. Using controlled anonymity: This general approach, also discussed in the next section, aims at restricting the users who can initially provide recommendations to the system. What is critical is that the user P has a reachable, fixed trustworthy real world existence that is identifiable by some attribute that is assumed to be non-changeable (such as an email, home address, IP number, telephone number). Using these and possibly other attributes, such as the "trust level" of that user, or whether or not the user is registered, allows the recommendation system to only accept

recommendations from "trusted" users and ignores the attempts by any other user to manipulate the system.

2. Using median filtering: This approach is a more technically oriented one. It involves replacing the sample mean by the sample median calculation of the MRE (Mean Rating Error).
3. Frequency filtering: This method is also a technically oriented one. It involves on estimating the average frequency of ratings provided by each buyers for a given seller. These filters can provide effective protection against ratings manipulation by insuring that the ratio of unfair raters in the MRE calculation set cannot be more than double the ratio of unfair raters in the total buyer population.

It can be observed from the discussion on security that system security is an essential pre-requisite for users to trust a recommendation system. If the system is vulnerable to attack, ratings and recommendations provided by the system can be misleading and lead to trust deterioration among the users of the system. However, as important as security is, it is not the only factor that influences user's trust in a recommendation system. Trust is a pre-requisite to the success and continued existence of any market and firm in general (Dellarocas, 2006). This is especially true in the online environment, which is time and space independent and can connect people from different times zones and different geographic locations as if they were neighbors. Studies have demonstrated that the more both sides of a transaction are separated in time and space, the greater the risk (Dellarocas, 2006). There is always the temptation by one side to change the agreed upon terms or try to sidestep certain rules that do not fit their needs. In traditional brick and mortar stores, trust is developed based on several cues, which can be rational or intuitive. For instance, trust can be built upon the appearance of a trading partner, their intellectual sophistication and knowledge, body language, the tone of voice and recommendations from friends or family. In the online environment, the case is much more complex. The underlying problem is in essence a problem of central authority and jurisdiction. Since the online medium spans multiple jurisdictions over multiple nations, determining who will enforce and ensure that "trusted" and agreed upon deals and contracts are respected and upheld by the law is a critical trust issue. This sophisticated situation is further complicated by the fact that trading partners can disappear overnight and appear with a new online identity at a different location (Dellarocas, 2006).

3.9. Trust and Trust building in Collaborative Filtering Systems

Trust is a major issue facing businesses in the modern economy, especially in the realm of E-commerce. Without a certain level of trust, no business transaction will successfully be completed, and no society will become civilized and flourish. Trust is a fluid expression that has several meanings and applications. For the purpose of this study, a very good definition of trust is provided by Dellarocas (2006), where trust is defined to be:

1. The individual is confronted with an ambiguous path, a path that can lead to an event perceived as beneficial or harmful
2. The individual perceives that the occurrence of benefit or harm is contingent on the behavior of another person

3. The individual perceives the strength of harm to be greater than the strength of benefit

The use of the word "perceives" in the above definition implies that trust is subjective. Psychologically, this means that the decisions made by individuals on the basis of trust are decisions made based on subjective views of the world. Building trust in the online world is much more difficult than in reality, for numerous factors which include the lack of face to face interaction, lack of legal jurisdiction over the online environment, the ease by which trade partners can "disappear" and "reappear" as a new online identity, the lack of trust people have in technology or a multitude of other factors. Trust building is a slow and difficult process that requires great care and attention to detail in its management. In order for recommendation systems and other similar systems, which require user input to function as anticipated, users must have a minimum degree of trust in that system and in the means by which that system exploit's the user's personal information (such as address, telephone number, credit cards etc.) This is true for any business or organization irrespective of the type of business that is employed. The delicate part of this trust relationship between customers/users and an organization is that users have no control over their own personal data once it is handed over to a firm (Miller et al, 2004). Users must entrust the organization to protect their personal information and to honor their privacy statements and privacy policies. However, in reality, not all firms are honorable, and privacy statements can be modified anytime to fit any vision the organization wishes, with the user completely helpless and unaware concerning the decisions regarding their own personal information. Concerning this situation, Miller et al (2004) have identified the following 5 potential risks facing users who wish to provide their personal information to firms online:

1. Deception by recipient: A website could intentionally seek to trick users into revealing personal and financial information, for the purpose of identity theft or other malicious goals.
2. Mission creep: This refers to a situation where the organization begins with a clear-cut defined use for personal information of its users, but gradually expands the original purpose over time with the ability to sell customer information to other firms to increase profit margins.
3. Accidental disclosure: this situation occurs when the user's personal information is made public by mistake over a website.
4. Disclosure by malicious intent: This situation describes exposure of the personal information of users by malicious means (such as hacking and theft.)
5. Subpoenas: Organizations are obliged by law to disclose user personal information when they are subpoenaed. This category can include the new trend of revealing customer information to the governments for reasons of national security.

The aforementioned risks and concerns occur after a user has shared personal information online with an organization. However, an important question to explore is what factors actually lead to the decision by users to share data online with firms initially, knowing the risks associated? It seems important to understand the user psychology in order to create systems that address the issues of concern to users. Miller et al (2004) provide the following 4 factors that their research has identified which explain the determinants for users to share data online with firms:

1. Whether or not the website shares information with another firm
2. Whether the personal information is used in a clear, identifiable way
3. The kind of information collected
4. The purpose for which the information is collected.

Hence, firms wishing to gather user personal information must first obtain the users trust. This can be partly accomplished by posting clear, concise and complete privacy statements and respecting the self-established rules. The above four concerns should be clearly addressed so users do not feel their data is being gathered for a malicious purpose. The purpose for gathering user information, explanations on why a certain kind of information is collected and a clear description of the purpose for which all the user information is collected should be sufficient to provide a basic level of trust between the user and an online firm.

From the above discussion, it appears that there exists several levels or layers of trust users become acquainted with. First of all, there is the highest level of trust, that between the user and the firm. This high level trust is important since studies have demonstrated that customer trust is positively correlated with a customer's intention to transact, purchase a product and return to the website (Pu and Chen, 2006). Secondly, there is trust between the user and the system being used. Thirdly, users need to trust the recommendations provided by recommendation systems themselves, and this can be facilitated by providing the proper feedback, and provide explanations and clarifications regarding the feedback. In their research on this issue, Pu and Chen (2006) emphasize that explanations can provide improved system transparency and thus increase the user acceptance and trust. Providing explanations is also thought to improve recommendation accuracy in some cases (Pu and Chen, 2006). It can be observed that providing explanations can elevate the user trust in a system, but this still does not address the issue of users trusting a recommendation by other anonymous users whose rating behavior and tastes are generally unknown. To address this issue, some concepts have been proposed in the literature, with the prevailing trend suggesting a recommendation system web architecture that provides the ability to define a "trust value" for each user P, usually a probability between 0-1 or a rating on a 1-5 Likert type scale, for each user P, to be provided by ratings from other users in the system whom the user had previous interaction with (O'Donovan and Smyth, 2006). Using this method, the "trust value" attribute can, for instance, be recorded as the total sum of successful/failures of all of user P's recommendations. This trust value can then be used to indicate to other users the trust "level" of P, in terms of recommendation quality and what other users think of P from their interaction experience. This method also requires analyzing the rating histories of each user to determine the predictive accuracy of the recommendations provided by that user. This approach has been successfully applied in e-commerce, with eBay.com being the most successful example. eBay encourages both parties involved in any transaction to provide feedback regarding each other after every transaction, and this information is made public to all users. This feedback consists of a numerical trust rating (-1, 0 or +1), along with short comments that other users can utilize to grasp an understanding of who the user is dealing with and what to expect from this user in terms of behavior and whether or not this user is to be trusted (Dellarocas, 2006). Although the research on trust in the online environment is still in infancy, this approach appears to be a step in the right direction.

Despite the extreme importance of trust and trust building in E-commerce and the online investment sector, this factor was not included in the model and is out of the scope of this research for several reasons. First of all, trust is a very broad topic which deserves its own research, especially since it encompasses the entire E-commerce realm and beyond. A more comprehensive E-commerce research would be required to accurately capture the user's feelings of trust especially in a complex domain such as online investment. Secondly, trust would be difficult to assess and measure using a prototype system. Since trust is a slow process that requires sustenance, a prototype system would not be able to capture this process nor will it be able to capture the investor's true feelings of trust. In order to more accurately capture this process, a longitudinal study would be required with commercial professional software that investors would be familiar with and feel they can trust in the sense that this software will not in any way manipulate or maliciously pass on their credentials to others. Finally, since the introduction of recommendation systems to the online brokerage domain has not been available commercially in the past, it appears rather difficult to expect users to trust a new technology the first time they use it and measure that trust during the brief exposure period they have with the prototype system.

3.10. Role of Explanations on Trust Building

As mentioned in the above discussion, trust building in the online environment is a slow and tedious process. Systems that employ collaborative filtering technology essentially aim at facilitating customer interaction, cutting down surfing time, and allowing customers to locate the product or service they need rapidly and with the least possible cognitive effort. **Table 6** provides an overview of the benefits investor obtain from going online. The general impression system designers and researchers have is that customers "like" recommendation systems and prefer to use them when they can. However, studies have shown that 21% of customers have negative reactions towards recommendation systems, which contradicts most of these rosy assumptions held by businesses (Wang and Benbasat, 2005). **Table 7** summarizes the main determinants which impact investor attitudes towards collaborative filtering .From the user's perspective, the apparent problem is not the actual collaborative filtering technology, but rather the lack of clarity and transparency as to why these systems provide these specific recommendations. Customers appreciate the role of collaborative filtering systems but they feel the current functionalities are insufficient.

1. Lower commissions, cited by 47% of respondents
2. Reduced cost of transactions
3. Not incurring extra costs for stop limit orders (35% of the respondents)
4. Convenience, availability of trading at any time
5. Perceived control over the system more than offline channels
6. Investment education

Table 6: Factors explaining why investors switch to online brokers

1.	Type of e-business website
2.	Product type being purchased
3.	Type of user
4.	Reason the item is being purchased
5.	Purchase for self or firm
6.	Whether subjective measures are relevant to the item.

Table 7: Determinants impacting user attitudes on Collaborative Filtering

In their study regarding the role of explanations in initial user trust building in recommendation systems, Wang and Benbasat (2005) have identified that the "trust" a user has in a recommendation system is actually composed of the three "belief components" of competence, benelovance and integrity of the system:

1. Competence: This belief holds that the user believes that the trustee (i.e. the recommendation system) has the ability, skills and expertise to act effectively.
2. Benelovance: This belief holds that the user believes that the system cares about the user and acts in the user's favor
3. Integrity: This belief holds that the user believes the recommendation system adheres to a set of rules and principles that the user identifies as acceptable.

The Theory of Social Responses to Computers (Reeves and Nass, 1996) states that people deal with computers as if they are social actors, and consequently people apply social rules to computers whether consciously or subconsciously. More than 30 additional empirical studies have demonstrated that regular users and even experts in the technology domain treat computers as if they are other individuals (i.e. humans) rather than a useful tool. This implies that users perceive the computer to have human characteristics, although in reality the computer does not posses such attributes. People tend to assign certain personality traits to computers, such as user-friendliness, helpfulness, annoyance or cruelness (Wang and Benbasat 2005). Upon the first interaction between a user and a recommendation system, we can comfortably assume that a user would be looking for some sort of cue in order to evaluate how trustworthy the system seems to be. The use of explanations in the recommendation system has been empirically proven to assist trust building in this critical initial phase of the user-system interaction. Explanations foster and encourage the trust building process, as the user can infer their own judgments and decisions based on the recommendation and explanations provided on one hand, while on the other hand these explanations, when provided in clear and concise manner, eliminate most doubts users have regarding the system and its transparency. Wang and Benbasat (2005) also identified that not all explanations are the same. Explanations in recommendation systems can be broken down into 'how' explanations, 'why' explanations and 'guidance' explanations, each best suited for a certain trust building purpose:

1. How Explanations: These explanations attempt to clarify to the user the logical reasoning involved in producing the final recommendation that the user currently sees. In other words, how explanations reveal the steps the recommendation systems utilizes to reach a recommendation to the user. How explanations aim at elevating user trust in the competence and benelovance of the system.
2. Why Explanations: These explanations serve a dual purpose. Firstly, they justify why the users are being asked certain questions and why certain details are required from the users (such as contact or financial information). Secondly, they reinforce

the justifications for the recommendations that are provided (for instance, no other similar products are in stock or this product is better because….). These explanations are extremely useful to identify the "intentions" of the system for the user in order to foster trust building. The use of why explanations are aimed to increase the user's trust in the system benelovance.

3. Guidance: Guidance explanations aims at assisting users who are unsure about what exactly they are seeking. Not all users are experts who know precisely what they want. Without a certain level of product knowledge, users and especially novice users might be unable to express their goals accurately to the recommendation system. This can result in users developing negative emotions towards the system and can negatively impact the entire trust building process. The role of guidance is to "enlighten" users in their decision making process and facilitate their trust in the system integrity. These explanations can provide users with suggested or informative information, leaving users to select their own course of action.

The how and why explanations are the main building blocks for the explanation interface. Despite it's utilization for a number of years, collaborative filtering systems nowadays still lack explanation interfaces in general and provide their recommendations without any descriptions or clarifications. Most recommendation systems still do not offer this services despite the fact that explanation facilities were rated as the 4th most important (out of 87) knowledge based system capabilities for users (Wang and Benbasat, 2005). It should be noted that trust does not fall on the shoulders of the recommendation system alone, but falls on the shoulders of the users as well in the form of the user's personality.

Different users exhibit different trust levels in real life and according to the Theory of Social Responses to Computers, different users will exhibit different trust levels in computer systems as well. Trust propensity is a personality trait that will impact the probability that a user will trust a certain entity. Wang and Benbasat (2005) have empirically demonstrated that trust propensity is indeed a factor that impacts the trust level users have in a recommendation system.

In terms of its operationalization and implementation, integrating an explanation interface into a recommendation system helps to open the "black box" of recommendations to users, providing them with justifications on why a particular recommendation is produced and the process that led to the presentation of that particular recommendation. Utilizing this approach makes users feel more involved in the searching process, and facilities user acceptance of the recommendation system. Moreover, users gradually become acquainted with the system as explanations provide a kind of "education" to the novice user. Herlocker et al (2000), in their detailed study of explanation interfaces and recommendation systems, concluded that the most preferred form of explanations for users is operationalized as a histogram of the neighbor's ratings regarding a particular item. Using a histogram provides the best explanation experience, based on empirical results, to users as it requires the least possible cognitive effort to examine and comprehend, namely a quick glance (which is a form of single, binary eye comparison). Hence it is recommended that implementation of explanation interfaces consist mainly of histograms and possibly other supporting forms of explanations.

3.11. Metrics used in Recommendation Systems

To better grasp and comprehend the various concepts related to recommendation systems and collaborative filtering research, it appears important to discuss and define the various metrics used for assessing and judging these systems. Most of these metrics, especially the more popular ones, are considered a standard in collaborative system research and they include:

1. Accuracy. To assess the accuracy of a recommendation, usually the MAE, (Mean Absolute Error) is used. MAE measures the difference between the actual and the predicted rating. This is the most commonly used measure to estimate the accuracy of a recommendation system (Mild and Natter (2002), Herlocker and Konstan 2002; Basilico, Justin 2004)
2. R squared: R^2 measures the correlation between the model forecast and the actual ratings. This measure can also be used for accuracy assessment. Other accuracy measures include Mean Squared Error and Normalized MAE (Herlocker et al, 2004)
3. Recall: Measures how often a list of recommendation contains an item that the user has actually rated. Calculated by generating a list of the top 10 recommended items for each user and testing whether or not an item left out of the user's training data is present on the top 10 list (Miller et al, 2004; Herlocker et al 2004)
4. Coverage: Measures the percentage of items for which a recommendation system can provide predictions.
5. Expected Rank utility: in many applications, correctly ranked items may be more important than prediction ratings for individual items. This metric was proposed initially in Breese et al (1998). This measures the expected utility of a proposed

ranking by scoring items in a recommendation list from top to bottom using an exponential discount factor (Basilico and Hoffman, 2004)

6. Pearson correlation coefficient. This is the most popular correlation coefficient metric in recommendation system research. It corresponds to an inner product between normalized rating vectors. This is the most commonly used measure to evaluate the "like-mindedness" between users. (Cheung, 2003)
7. Density reflects both the overall size of the recommender system space and the degree to which users have explored it (Herlocker et al , 2004)

4. Online Investment

4.1. What users expect

In the modern era of user-centered design, understanding the needs of the customers and users of a system is a critical design requirement. The more aligned the design is with user's expectations and needs, the higher the probability of user adoption and satisfaction, which in business terms can translate into increased profits, higher customer loyalty and increased customer retention rates (Vucetic and Obradovic, 2005). This fact is especially true in the online brokerage environment, where increased customer satisfaction and trust in the system is critical for further profit. Users tend to evaluate and assess their satisfaction of the services provided by online brokers through several cues which include perceived operational competence of the system, perceived trustworthiness of the online broker, and the perceived cost of trading to execute orders (Balasubramanian et al 2003).

The introduction of the Internet and its worldwide adoption has complicated the task of system designers. In the past, designing software and information systems was generally aimed at a group of users who share many similarities, and who were more professional in their usage. Nowadays, more and more Internet and E-commerce customers are less professional users and come from a broad spectrum of backgrounds, cultures, religions and use different languages in their daily lives. Designing for universal usability, while maintaining user-friendliness and ease of use is a constant challenge for system designers. Commenting on this issue in the realm of online investing, Balasubramanian et al (2003) concluded that most complaints received from users regard operational factors in particular. The main complaints were user's inability to access websites, delays in order processing, poor contingency planning, unsatisfactory timeliness of trade, poor quality of provided information, the need for online discussion forums and privacy issues relating to their personal data (Balasubramanian et al 2003).

One of the most critical user needs concerning e-commerce recommendation systems is the availability of the explanation interface discussed in details throughout the previous sections. Explanation interfaces assist putting users at ease and facilitate the trust building process (Pu and Chen, 2006). It is important to point out that Herlocker et al (2004) concluded that the preferred explanation interface for users is based on how a user's neighbor rated an item and similarity of the recommended item to other items the user has rated highly. Another important issue for users is the length of time required to complete a purchase. Designers of new e-commerce systems should attempt to reduce the cognitive and physical effort necessary for purchasing to only one-click purchases for standard and repeat purchases. **Table 8** summarizes the main design guidelines proposed by various researchers. Moreover, the payment process and verification of personal and financial information should be quick, easy and trustworthy (Bellman et al, 1999). Users also like to feel in control, especially in the personal financial management domain. Users want control of their investment decisions, personal information, and on deciding which information to save and which information to discard (Alpert et al, 2003). Users have also demonstrated that they do not appreciate implicit gathering of user behavior, and this should be avoided when possible. Alpert et al (2003) concluded in their study on user attitudes regarding user adaptive E-commerce websites that users had a negative reaction to implicit information gathering, which can in turn negatively impact their perceived trust and competence of the system. Implicit information gathering, when viewed in this sense, is potentially a hindering force promoting user distrust in the intentions of the system, and its use should be done with extreme caution or avoided altogether.

1. Provide explanations to communicate the recommendations to users. This can be operationalized as histograms of the ratings of similar users and can include interfaces explaining why this recommendation has been provided and/or how it has been provided (Herlocker et al, 2000)
2. Design for universal usability and user-centered design
3. Simple, user friendly and easy to use interface (Davis 1989)
4. Sites should make it more convenient to buy standard or repeat purchase items such as one click purchase. (BellMan et al ,1999)
5. Payment process should be easy and secure (BellMan et al ,1999)
6. Categorize recommendations according to their similar tradeoff properties relative to the top candidate. This approach can improve the accuracy of the user's decision by up to 57%. (O'Donovan and Smyth, 2006)
7. Propose improvements and compromises in the category title using conventional language (while keeping the number of trade-off attributes below 5 to avoid information overload (O'Donovan and Smyth, 2006)
8. Eliminate dominated categories and diversify the categories according to their titles and contained recommendations (O'Donovan and Smyth, 2006)
9. Include actual products (stocks) in the recommended category (O'Donovan and Smyth, 2006)
10. Rank recommendations within each category by "exchange rate" relative to the top candidate (pros and cons) rather than similarity measure (O'Donovan and Smyth, 2006)
11. Avoid implicit information gathering when possible, Alpert et al (2003) concluded in their study on user attitudes regarding user adaptive E-commerce websites that users had a negative reaction to implicit information gathering, which can in turn negatively impact their perceived trust and competence of the system .
12. Users want to be able to view, modify, delete and add personal information at any time (Alpert et al, 2003)
13. Users want to feel "in control" but they should be warned of over confidence especially in the online medium
14. Users want explicit control over the ability to save data

Table 8: Design Guidelines for E-Commerce Systems

Collaborative filtering systems are relatively a new technology in the online environment. Many users are still not familiar with the technology and what the technology can actually achieve. Naturally, different users hold a variety of perspectives regarding collaborative filtering, with some promoting positive views while others demonstrate negative views. In their study on user attitudes regarding user adaptive websites, which was performed in a more general context than recommendation systems, Alpert et al (2003) have identified the following factors that impact those user attitudes:

1. The type of e-business website
2. Type of product being purchased
3. The type of user
4. Reason the item is being purchased
5. Purchase type (for the self or for work)
6. Whether or not subjective measures are relevant to the time.

Regarding the design of collaborative filtering systems in particular, Pu and Chen (2006) compiled a very useful set of 5 core guidelines and suggestions for designers of recommendation systems based on responses from their empirical study. When providing recommendations on a selected item, the recommendation system should categorize the remaining recommendations according to their similar tradeoff properties relative to the top candidate. Tradeoff support is estimated to improve the user's decision accuracy by up to 57%. The recommendation system should also propose improvements and compromises in the category title using conversational language, while simultaneously maintaining the number of tradeoff attributes below 5 to reduce information overload and short term memory overload for the user. Using natural language helps the users to feel at ease; with users preferring phrases such as "these stocks have a lower price and lower risk probability" over phrases such as "cheaper but less risky stocks". Moreover, the recommendation system should diversify the categories in terms of titles and contained recommendations, while focusing on eliminating "dominated" categories. Dominated categories refer to categories that are very similar to other categories in terms of their tradeoff properties. Dominated categories do not provide substantial recommendation influence to users and should thus be eliminated. Recommendation systems should also include actual products or services in the recommendation category, and not just the titles or names. Customers were found to achieve their goal faster with an average of about 6 products recommended in each category. Finally, recommendation systems should rank the recommendations within each category by their "exchange rate" or tradeoff properties relative to the top candidate, emphasizing the potential gains and losses for that product when compared to the top candidate (for instance, this stock is more expensive but less risky than stock A). Finally, in their study on the implications of online investment, Konana et al (2000) identified that investors desire a transparent window in the interface to view the actual flow of orders, the actual time of the execution and the commission fee structure at various points in the trading process in order to make more sound investment decisions. It is suggested that designers of electronic trading systems adhere to the above mentioned guidelines and incorporate them into their design process. Although the suggestions appear initially as less important than other design issues, they are based on relatively new and solid empirical studies that utilize consumer feedback, and these guidelines could be the factor that provides competitive advantage for the firms employing them over firms that choose to ignore these suggestions.

4.2. Prototype Introduction

Herlocker et al (2000) suggested that for a recommendation system to achieve the optimum performance, it appears wise to specifically identify the tasks for which the system is required for. Moreover, the researchers also suggest combining both collaborative filtering and content filtering approaches to provide the best possible recommendation quality and accuracy. For the sake of this research, only a prototype is necessary to perform empirical testing of the hypotheses. The design of commercial recommendation system software is a more complex process that has more rigorous requirements and needs to provide more advanced functions. Based on this advice, the following tasks have been identified as important for the prototype recommendation system under discussion:

1. Provide a database with data regarding 50 stocks that is accessible via web browser
2. Combine collaborative and content filtering approaches
3. No real time data connectivity is required, all calculations can be done on past data obtained from financial websites for the sake of this experiment
4. Provide a detailed questionnaire to users upon their first visit in order to categorize users into classes/neighborhoods based on their risk adverseness
5. Provide a user friendly, easy to use interface that mimics basic functionality in commercially available electronic trading products
6. Allow feedback from users regarding the system. Feedback should be stored in a separate database for later analysis

5. Research Design

5.1.Research Model and Hypotheses

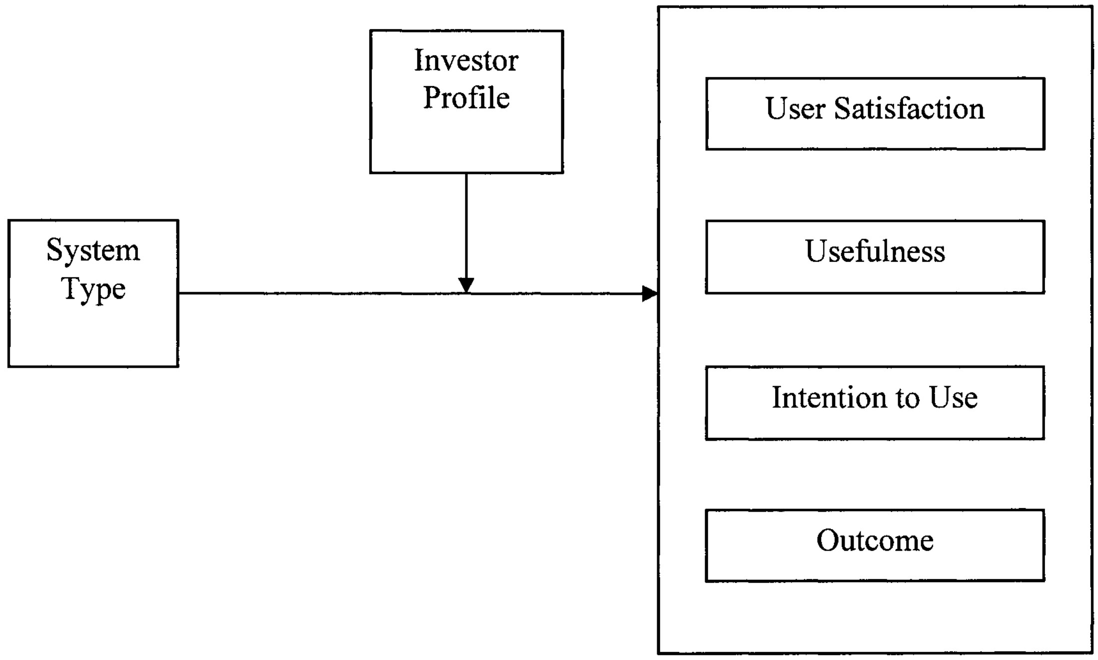


Figure 2: Proposed Model

The proposed model incorporates several variables and their relationship to user satisfaction, user perception of the usefulness of the system, the user's intention to use the system and the outcome, measured using the BETA, also known as the stock's systematic risk. The model is based on the concept of design research and not behavioral research. Behavioral research would likely to include links between user satisfaction, usefulness and intention to use. However, for the purpose of this research those links were not investigated. The model is partly based on the logic of the Delone and Mclean (1992) IS Success Model, which states that the information quality and system quality have a direct positive correlation on user satisfaction and usage of the system. In this model, system type can be viewed to represent information quality since both systems were identical except for the critical difference of recommendations (information) provided by the second, advanced system. However, this model is not based on Delone and Mclean but follows similar logic regarding the information system. The model consists of 6 main factors which are explained in more details below:

1. System Type: System type is used to refer to the version of the system being used. This is a binary variable with 0 representing the Basic System which provides no recommendations to the investor, and 1 representing the Advanced, recommendation system.
2. Investor Profile: Investor profile is a variable used to reflect the risk attitude and the investment "know-how", and how these impact the investor's perception of usefulness, intention to use the system, the outcome and user satisfaction. The Investor Profile consists of Risk Attitude, Investor Knowledge, which measures the overall investment information and knowledge the participant holds and Investor Experience, which measures the years of experience the participant has with investment.
3. User Satisfaction: This variable represents the overall user satisfaction of the investor with regards to the prototype. It is measured by capturing the responses of the investor to the user satisfaction questionnaire and calculating a satisfaction score.
4. Usefulness: This variable is used to represent how useful the investor perceives the system to be. It is measured by capturing the responses to answers in the questionnaire and calculating a usefulness score.
5. Intention to Use: This variable represents the degree to which the investor intends to use the system once it is commercially available. This variable is measured by capturing the investor responses to the corresponding questions in the questionnaire and calculating an intention to use score.
6. Outcome: This variable is used to represent the overall BETA of the investor's portfolio after exploring the prototype and purchasing the stocks of interest. The

higher the BETA, the more risky the portfolio. This variable is calculated as an average of the BETA of the purchased stocks.

The rationale behind the model is that the type of system being used, along with the investor's profile which is unique to each investor, have a direct impact on the user satisfaction, user perception of usefulness of the system and the intention to use the system in the near future.

The essence of this empirical study is to validate the following research question:

What is the effect of introducing collaborative filtering recommendation systems on user satisfaction, user's impression of usefulness of the system and the intention to use the system in the domain of online brokerage?

The experiment was designed to assess the following hypotheses:

H1: Users of the collaborative filtering advanced system will have a significantly higher user satisfaction score than users of the basic system.

This hypothesis follows the rationale in the Delone and Mclean IS success model. It generally states that since the collaborative filtering system will provide meaningful information in the form of stock recommendations based on the behavior of similar users, investors using the system will be more satisfied with the collaborative filtering system over the basic, non-collaborative filtering system which provides no extra information besides the general financial information such as the stock's price, the BETA of the stock, an overview of the stock's history and similar data.

H2: Users of the collaborative filtering advanced system will perceive it as more useful than users of the basic system.

When investors are given the chance of obtaining valuable recommendations based on the behavior of similar investors, it would appear logical that they would find this information valuable since it has a direct meaning and significance to them as the information is customized and adjusted based on their risk tolerance, and not just generically provided for all investors.

H3: Users of the collaborative filtering advanced system have a higher intention to use the system once it is available than users of the basic system.

If investors perceive that an information system provides them with valuable information and adds value to their investment decisions by providing customized recommendations, they would be more likely to have a higher intention to use that system over another information system which does provide meaningful information tailored to their specific needs and rather treats all investors in the same manner.

H4: Users who are risk takers will be more likely to select riskier portfolios (those with higher BETA) than investors who are conservative, those with limited or no investment knowledge and experience.

Risk loving investors, on average, have a higher investment knowledge and experience. These investors would more likely aim for riskier portfolios that have a higher return when they have the opportunity. On the other hand, more conservative investors would not take such risks and would rather invest in a more predictable portfolio which would provide a lower return in exchange. The measure of risk for the portfolio utilized in this study is BETA, a financial measure which represents the stock's systematic risk.

H5: Users with less investment knowledge and experience will have a higher user satisfaction, usefulness and intention to use scores with the advanced recommendation system than more experienced and knowledgeable users.

This hypothesis follows the rationale that less experience and knowledgeable investors will perceive the recommendations provided by the system as a useful guideline, especially since they are new to this financial domain. This in turn will translate into higher satisfaction, usefulness and intention to use scores. On the other hand, more experience and knowledgeable investors are assumed to have previous experience with financial decision making and would perceive the impact of recommendations as less effective and useful than the novice investors, which in turn would translate into lower satisfaction, usefulness and intention to use scores.

For the purpose of this experiment, the following three determinants were measured:

1. User Satisfaction with the recommendation system
2. Usefulness of the recommendation systems
3. Intention to use the recommendation system once it is commercially available

User satisfaction, usefulness and intention to use were captured via a 12 question survey using a Likert Scale of 1-7 for measuring each item, where 1 represents one extreme (such as extremely dissatisfied) and 7 representing the opposite extreme (such as extremely satisfied). The questionnaire is available in the **Appendix**. Results of the questionnaire were analyzed using SPSS statistical software.

5.2. Experimental Setup

The online experimental prototype was developed using Macromedia Coldfusion, Dreamweaver MX7 and Microsoft Access. The rational behind the prototype was to develop two identical versions of the system but with only one critical distinction: the first basic system provided no recommendations to the users throughout their exploration of the available stocks while the second advanced version included a collaborative filtering code that allowed the generation of dynamic recommendations, using the Jaccard Index (an index which compares the similarity of two objects, which in this specific case are stocks in the investor's portfolio) based on similar simulated portfolios already stored in the investor database. The recommendations were calculated and implemented based on the BETA of the stocks, or in other words the systematic risk of the stock. The higher the BETA, the more likely the stock will not be correlated to the market risk and the riskier it is. Although there are other more comprehensive measures used to assess the risk of a stock in the financial world, BETA is often used as well and provides an indication regarding the stock with respect to the market. Initially, investors were provided a rough recommendation based on content filtering with respect to their risk assessment score from the first questionnaire.

5.3. Website Development

The prototype essentially consisted of the following web pages (in sequence) programmed using Macromedia Dreamweaver MX:

1. Login screen to capture investor information and demographics.
2. Risk assessment online questionnaire to measure the risk attitude of the investor and provide feedback.
3. Instructions on what to expect and how to proceed with the experiment.
4. Stock exchange system for trading stocks. Since this prototype was developed for research purposes, the system only showed 50 pre-selected stocks and provides information stored in the database and not in real time. The system allowed purchasing stocks, removing stocks, clearing the portfolio of all stocks as well as exploring various information regarding the stocks and the definitions of the terminology used (such as BETA).
5. User satisfaction questionnaire to assess the satisfaction level of the investor after using the system.

The prototype also included several pages of instructions before and prior to the main sections above illustrating what the participants need to know in order to proceed to the following step. Once the systems were functional on a server, emails were sent to the participants asking them to participate in the study. The emails included a link to the website. The experiment was conducted online using a convenience, non-probability participant selection method. Participants would then click on the link, enter their demographic and login information and then proceed with the experiment. Participant data and answers were captured and inserted as entries into a database. This information was then used to analyze the user responses to questionnaire items as well as the purchased stocks of interest.

6. Research Findings

6.1.Results

The experimental prototype involved 70 respondents who were selected based on their acquaintance with the basic understanding of online investment and its terminology by the use of a convenience, non-probability sampling method. The sample was selected using the Snowballing method, whereby each participant was sent an email with the link of the online prototype system and asked to forward the link to any acquaintances which would like to participate and who are familiar with online investment. This sampling method is preferred when groups of participants are difficult to identify, such as in this case. There is no clear cut description on how to divide the participants into groups, and therefore the applicability of non-probability sampling appears justified since any individual can be an investor (WebSM, 2006). Initially, 90 participants were asked to join via email, of which 30 responded with a response rate of (33.3%) while the remaining participants were the result of Snowballing (40 participants representing 44.4% of the participants). Of the 70 participants, 25 (35.71%) were female while 45 (64.29%) were male. **Figures 3a**, **3b** and **3c** provide an overview of the breakdown of the respondents by age, investment experience and investment knowledge.

Figure 3a: Distribution of respondents by age

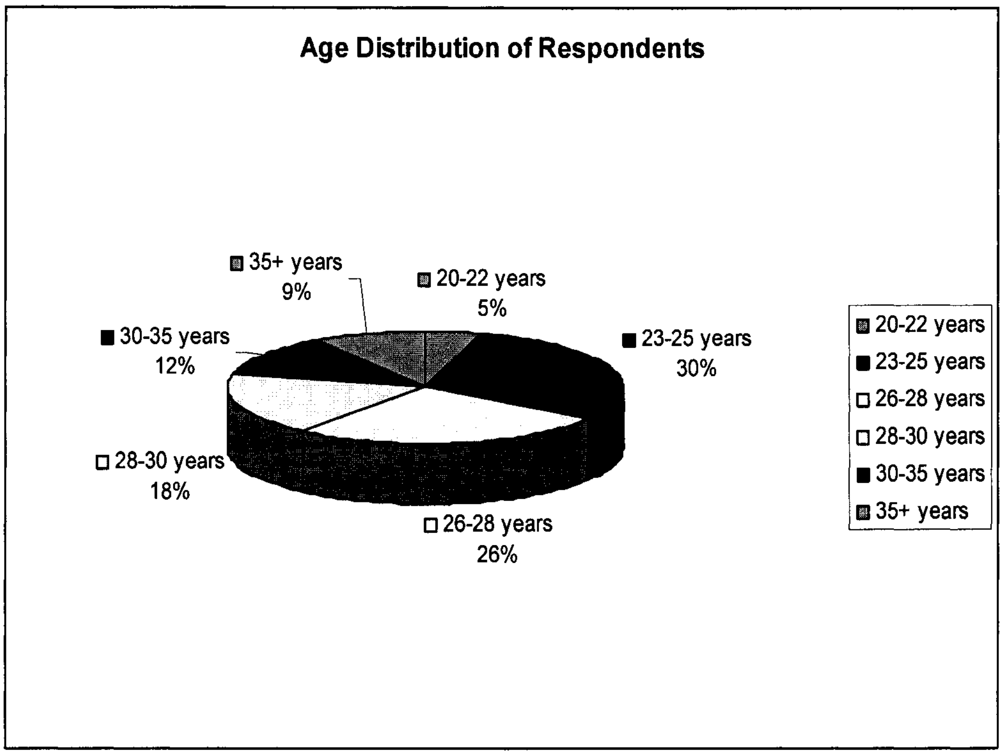


Figure 3b: Distribution of respondents by investment experience

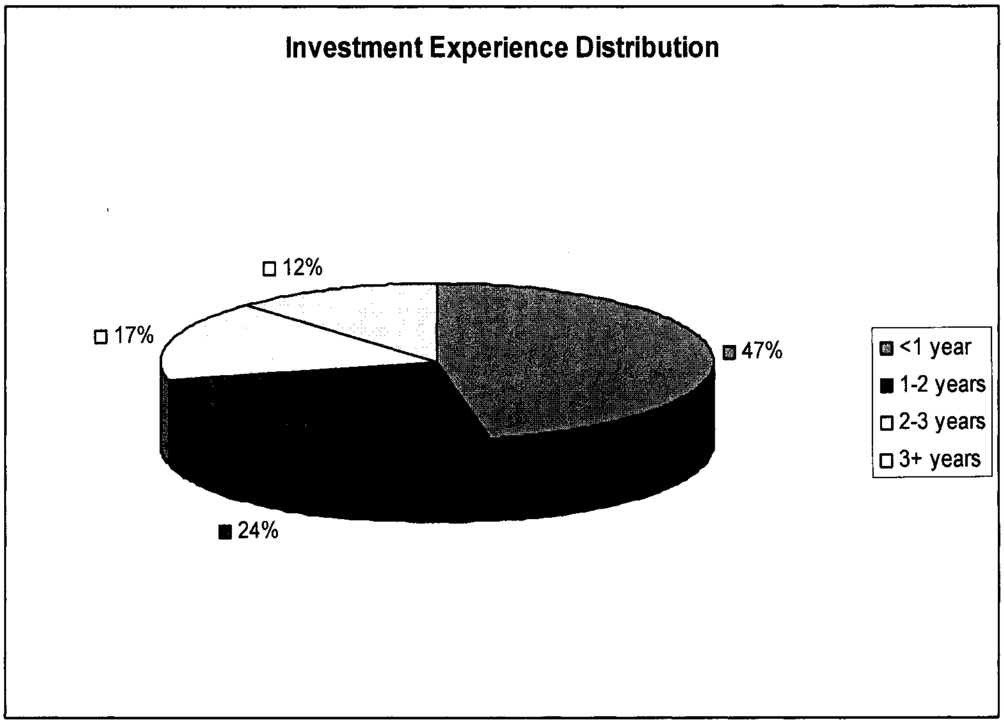
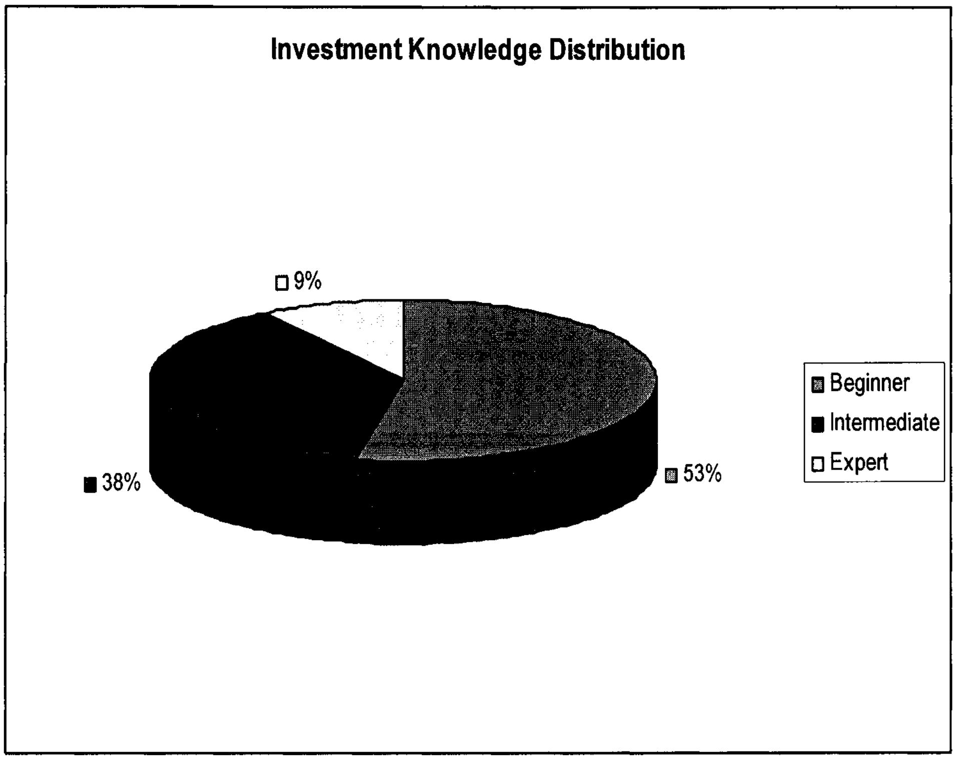


Figure 3c: Distribution of respondents by investment knowledge



The participants had an age range of 20-35+, with the average range being 26-28. The investors had an investment knowledge ranging from beginner to expert, with the average investment knowledge being intermediate. The investors also had an investment experience ranging from less than 1 year to 3+ years investment experience, with the average investment experience being 1-2 years. The Risk Scores for the participants ranged from 15-37, with the average Risk Score being 25, indicating a Semi-Conservative risk attitude on average for the participants.

Exploratory factor analysis was conducted to evaluate convergent and divergent validities of the constructs. Findings also indicate a solid coefficient of reliability for each of the factors being measured. **Table 9** summarizes the Cronbach's Alpha for each of the factors being explored. User satisfaction was found to have a coefficient of consistency of 0.88, usefulness was computed to have a coefficient of consistency of 0.83 and the intention to use the system was found to have a coefficient of consistency of 0.7.

Factor	**Cronbach's Alpha**
User Satisfaction	**0.88**
Usefulness	**0.83**
Intention to Use	**0.7**

Table 9: Cronbach's Alpha

Based on the findings, H1 was supported at $p<0.05$ level. It has been confirmed that users of the more advanced system which proposes recommendations to the investors have a higher user satisfaction score than investors using the basic system which does not provide any advice or recommendations. H2 was not statistically significant at $p<0.1$ and hence is not fully supported. However, H2 seems positive and in the correct direction thus it is possible under more stringent conditions and with a larger data set that H2 can potentially be confirmed. H3 was statistically significant at the $p<0.1$ level and thus H3 is supported and confirmed by the findings. H4 was found to be in the correct direction, similar to H2, but was not significant at $p< 0.1$ although it was very close to being significant (p-value=0.12). There is some evidence that more experienced and knowledgeable users with tend to select riskier portfolios. With more rigorous testing and a larger data set this hypothesis could prove significant. Finally, H5 was not significant at $p<0.1$ and was not supported by the findings. **Tables 10a** and **10b** below summarize the hypothesis testing. There were no statistical differences found between the two groups of participants. See the **Appendix** for the details and statistical p values.

Hypothesis	Mean Collaborative	Mean Non-Collaborative	P-Value	Status
H1**	**5.37**	**4.47**	**0.035**	**Supported**
H2	**5.95**	**5.76**	**0.268**	**Not Supported**
H3*	**5.19**	**4.46**	**0.012**	**Supported**

Table 10a Hypothesis P values

Note:

*: Supported at $p<0.1$

**: Supported at $p<0.05$

With regards to the moderator hypotheses H4 and H5, they were analyzed using the following group of equations:

H4: → BETA = A + B(Risk) + C(System) + D(Risk * System)

H5: →Y = A + B(System) + C(Investor Profile) + D(System * Investor Profile)

Investor Profile is replaced once with satisfaction, once with usefulness and once with intention to use. The coefficient of interest is the interaction variable, identified as D.

The p values and coefficients from the above equations are provided in Table 10b below:

Hypothesis	P-Value	Coefficient	Status
H4	**0.12**	**D=0.031**	**Not Supported**
H5	**0.206, 0.41, 0.13**	**D=-0.03, -0.06, -0.35**	**Not Supported**

Table 10b: Hypotheses Coefficients and P values

6.2 Discussion

Overall the findings support the notion that investors using the Basic System (System 1) had on average a lower user satisfaction score than users of the Advanced System (System 2) and thus users of the recommendation system were found to have a higher satisfaction score, hence H1 was supported at the $p<0.05$ level. On average, the user satisfaction of System 1 was 63.9% while the average user satisfaction for System 2 was 76.7%, resulting in users of System 2 having a 13% increase in user satisfaction over users of System 1. This difference can be attributed to the collaborative filtering recommendation system since all other components of the two systems were identical.

With regards to H2, users of System 1 had an average Usefulness of 82.3%, while users of System 2 reported an average Usefulness of 85.1%. Although the difference is positive and in the correct direction, the difference is not statistically significant and hence H2 is not confirmed. The difference of 2.8% usefulness score in favor of the advanced system can perhaps be improved if tested under a different set of conditions and with a larger dataset. Since the only difference between the two near identical systems was the collaborative filtering component, it appears this component had an impact on the investor's perception of usefulness although the finding was not statistically significant.

H3 was supported at the $p<0.1$ level. Users of System 1 reported, on average, a score of 63.7% with respect to their intention to use the system once it is available. On the other hand, users of System 2 reported a score of 74.1% intention to use the system once it is available. The 10.4% difference is positive and in the correct direction. This supports the hypothesis that users were more satisfied with the system and would like to use it again when they have the opportunity to do so. Since it was confirmed that users of System 2 had a higher user satisfaction score, it makes logical sense to expect them to intend to use the system once it is commercially as a comprehensive software package. The fact that usefulness was not found to be significant while intention to use was significant follows the logic found in the IS Success Model by Delone and Mclean, which states that user satisfaction with an information system and the use of that information system are very closely related and reciprocally interdependent (Delone and Mclean, 1992). However, when viewed from the standpoint of the Technology Acceptance Model (TAM), it would not appear logical that usefulness was not supported while intention to use was supported, since TAM and the online extended modifications of TAM have empirically proven that usefulness is positively correlated with intention to use while user satisfaction does not directly impact intention to use (Davis, 1989). Nevertheless, since this research follows in the footsteps of design research more than behavioral research, the significance of the findings is the actual user response to the system and their reactions to using the system.

H4 was not supported. However H4 was very close to obtaining statistical significance since its p-value was $p=0.12$. There is some evidence that more experienced and more knowledgeable investors tended to select portfolios with higher BETA values (used as a risk indication in this study) which translate into higher stocks selected in their portfolio. The results are in the correct direction and are very close to being significant at the $p<0.1$ level. However, the findings are not conclusive with regards to this hypothesis.

Perhaps with a larger sample size or a more comprehensive measure of portfolio risk this hypothesis could be supported. Since BETA measures the market risk of a portfolio, a more comprehensive risk construct could perhaps obtain a more statistically significant result.

H5 was not supported. The hypothesis stated that users new to the investment world would perceive the collaborative filtering system to be more useful to them (over experienced investors) since it can help guide their decision making resulting in a higher user satisfaction, usefulness and intention to use scores. On the other hand, more experienced users will not require the use of such services or aid since they have more trust and confidence in their abilities and knowledge. However this was not confirmed by the findings. On the contrary, the p-values indicate that there is a difference and it is in the opposite direction. The difference was not found to be statistically significant. It appears more experience and knowledgeable investors perceived the recommendation system as more important and useful to them than novice investors. Since collaborative filtering recommendation systems are based on the behaviors of other similar people, investors with limited or no experience could have perceived the recommendations to be based on other, similar inexperienced users, resulting in a lack of trust in the recommendations. Moreover, it could also be the case that more advanced users have higher expectations and thus perceive the idea of having a recommendation system based on the decisions of other, similar minded experienced investors to be useful to their decision making, especially if they are in doubt regarding some of their investment decisions.

To capture accurately the results of the questionnaire, 4 out of the 12 items were dropped resulting in 8 final questions in the questionnaire. However, it was tested and confirmed that the meaning of the constructs were captured and validated.

Overall the findings indicate that investors prefer the collaborative filtering recommendation system. Investors were more satisfied with using the system and intended to use the system again once it is available. This indicates recommendations provide users with important information, feedback and guidance.

6.3. Limitations

There were four main limitations with this research. The first encountered limitation occurred when recommendations were required to be generated for the participants dynamically. Both System 1 and System 2 of the research were performed simultaneously to avoid any potential errors that could arise from performing the research at different time intervals due to any possible changes or biases during that time frame. Since both systems were online and tested, simultaneously, there was no data to use in order to generate recommendations. The solution to this issue was to simulate 100 portfolios to use in the database as a basis for recommendation generation. The Risk Assessment Questionnaire divides investors into 4 main groups (classes) which include Aggressive, Semi-Aggressive, Semi-Conservative and Conservative depending on their Risk Score. The 100 simulated portfolios were broken down equally among those 4 groups, with 25 portfolios in each group. Within each group, the portfolio was prepared by randomly selecting stocks which belong to that risk group using the BETA of the stocks as an indicator of how risky the stock is. Since BETA measures the stock's systematic risk, it directly relates to the market. A higher BETA value suggests a riskier stock which does not follow the market trends. Using this technique, 25 randomly generated portfolios were available for each risk group and can be used to generate recommendations. Once investors began selecting stocks, the collaborative filtering code would match the selected stocks in the portfolio dynamically to the closest simulated portfolio in the database and offer the remaining stocks (those the investor has not selected) as recommendations.

The second limitation in the study was the assumption of equal weights for the stocks. In reality, portfolios consist of a variety of stocks (and possible other securities such as bonds and options) and different quantities of those stocks. The portfolio's risk is assessed and calculated with the weights of the stocks, which are derived from the quantity of the stocks relative to the portfolio, in mind. However, since this research was not aiming at obtaining the highest accuracy portfolio risk calculation but rather intended to measure the user perception of the concept of providing recommendations to assist them in making stock purchases, weights were not taken into consideration and the prototype did not allow the investor to purchases quantities of stocks. Investors could only select 1 from each stock they added to their portfolio.

The third limitation in the study is generalizability. Although the findings confirm H1 and H3, caution should be taken when generalizing the findings to other domains besides online stock trading. It was confirmed that investors in general would be more satisfied using online investment with some sort of a recommendation system. However, to generalize the findings of this study, further research would be required with a larger sample size, using different settings and a wider geographical dispersion to determine which domains of online investment require recommendations the most and what type of recommendations the investors would like to see. The fourth and final limitation in this research is the use of web experiments as opposed to a controlled lab experiment. Lab experiments have the ability to be almost fully regulated and controlled, removing any potential moderator or mediation effects and having the experiment conducted under the supervision of the researcher. However, in real life, investors would not perform their investment tasks inside controlled labs but rather from their home or office and thus the use of web experiments for this specific task appears justified as it captures the investor attitude and satisfaction in the same setting they would use a commercial investment software package. Research has demonstrated that web experiments are becoming more frequently used, and they provide several benefits including a solid internal validity, reliable external validity, lower financial costs and a much faster response time (Wade and Tingling, 2005).

6.4 Conclusion and Implications

Online investment is the wave of the future in the investment world. Gradually more and more people are switching to the Internet to satisfy their investment needs. As the Internet proliferates globally and the online community grows, more people will prefer to use the comfort, safety, speed and low prices of the Internet broker over full service brokers. Despite the numerous benefits provided with online investment, it can be a double edged weapon. Investors of the online community will not have easy access to professional advice and recommendations like those in the offline community. This, in turn, can negatively impact their performance and can lead to lower trading profits. One possible solution which is rather novel in its approach is to introduce a form of recommendation system, similar to that available for books and other products on Amazon.com, to their online investment needs.

The purpose of this study was to introduce recommendation systems to the online investment domain. The rationale was to statistically test the integration of a collaborative filtering recommendation system to a prototype stock exchange system where investors using the system can dynamically receive recommendations based on the stocks they purchase. The experiment involved creating two identical websites which mimic the main functionalities available in the stock exchange software currently in the market with one fundamental difference: One system was coded with a collaborative filtering recommendation system while the other system offered no recommendations or advice whatsoever. The study involved 70 participants from various age groups and backgrounds. Findings indicate that investors using the recommendation system had a higher user satisfaction (13% difference) than those who did not (H1 confirmed). They found the recommendation system to be informative and useful to their needs. Moreover, investors indicated that they had a higher intention to use to the system if it had a recommendation system than a regular system with no recommendations (H3 confirmed). Finally, users of the recommendation system reported a higher usefulness than users of the regular system. Although the difference is positive and in the correct direction, it was not statistically significant and hence H2 is not supported with these findings. H4 was not confirmed but was very close at being statistically significant. H4 stated that investors with more investment knowledge and experience will tend to select riskier portfolios (portfolios with higher BETA values). H5 was not supported and rather appeared in the opposite direction to what was hypothesized. H5 stated that investors with limited or no investment knowledge and experience would appreciate the recommendation system more and report higher user satisfaction, usefulness and intention to use scores. Contrary to expectations, it was found that investors with more investment knowledge and experience reported higher scores. It appears the more knowledgeable and experienced an investor is, the more the investor appreciates the role of the recommendation system.

Regarding the business implications of the findings, they generally can be summarized as follows:

1- Online investment offers a variety of benefits for both the investors and the firms using online brokerage. Benefits for investors include (but are not limited to) availability of free advice and information on the Internet, lower trade commissions, faster trade execution, and the ability to utilize the power of recommendation services when applied to the online investment sector, which is the main aim of this study. For firms, the advantages include personalized and targeted marketing and promotion, higher customer satisfaction due to convenience and lower costs, increased profits, higher customer loyalty and increased customer retention and return rates

2- It was confirmed that using collaborative filtering for the purpose of generating recommendations to investors regarding which stocks to purchase based on their risk attitude leads to higher user satisfaction with the system. Although generalizing the results beyond the scope of this study should be done with caution, it does provide empirical evidence for the brokerage industry that software packages which can provide personalized recommendations in a manner similar to movie and book recommendations provided by firms such as **www.Amazon.com** on their websites are desirable by investors and in fact lead to higher user satisfaction ratings. Moreover, investors perceived the recommendation system to be helpful and demonstrated their intention to use the system once it is commercially available.

3- Practitioners should open the door to incorporating collaborative filtering into the online investment domain and expand the use beyond stock exchange to perhaps all securities and even beyond. This domain appears very broad and financial information is extremely difficult to filter and grasp on the Internet. Since research has demonstrated that investors nowadays are more time starved, it would appear logical to incorporate newer technologies that perform collaborative filtering for their clients, perhaps with the use of agent-based techniques, to simplify their investment needs and help guide investors who desire recommendations and assistance.

4- Online investment firms which chose to begin incorporating collaborative filtering technology in the online investment domain need to take caution on several technical details with regards to the quality and validity of the recommendations. One important distinction, for example, between book or movies recommendations on one hand and stock recommendations on the other hand is the variability of the product attributes. For instance, books and movies do not have a large variability in their price, but on the other hand stocks can be extremely variable with regards to their price, their level of risk and their return rates. Moreover, stocks also have a very rapid rate of change with regards to its statistics and information, and this must be accounted for when designing a collaborative filtering system for stocks. Out-dated stock information can lead to poor or inaccurate recommendations which can consequently have large impacts on the investment decisions and behavior of many investors. Stock price volatility is an extremely important topic in the design of recommendation systems for online investment. It should be taken into account and requires future research to determine the impacts it has on the collaborative filtering recommendation systems for online investment.

6.5 Future Research

There are several potential avenues for future research within the realm of applying collaborative and content filtering methods to the online investment domain. This experiment only scratched the surface of such a complex domain, providing insight that investors would like to see recommendations developed by collaborative filtering to assist them in their investment needs. Future research can explore the application of such technology to investment products and packages currently in use by investors, to provide evidence of whether or not real life investors would benefit from such a technology. Other potential research avenues can expand the concept of recommendations based on collaborative filtering along to combine other technologies such as user ratings, real time market indices and content filtering technologies in order to provide investors with the most accurate and up to date recommendations possible. Furthermore, it is also possible to research the application of collaborative filtering to other domains besides stock exchange systems to also include currency exchange system and systems that trade in securities in general Collaborative filtering systems are currently used by Amazon.com and other websites to provide collaborative recommendations for books, movies, video games and other products. However, the results of this experiment support the notion that expanding this kind of technology beyond the simple market of books and movies is possible and even desirable by customers.

Other future research directions can attempt to explore the concepts introduced by the unsupported hypotheses, which sometimes can prove to be insightful. On one hand, H5 which was not supported and rather which appears to be in the opposite direction is an excellent candidate for future research. H5 stated that novice investors would appreciate and be satisfied with the collaborative filtering system more than expert investors. Results indicate that on the contrary, more experienced investors actually reported higher satisfaction scores and seemed to appreciate the technology more. Future research can explore what factors lead more experienced investors to appreciate the collaborative filtering system more than the new investors. Also, H4, another unsupported hypothesis, stated that investors would find the collaborative filtering system useful. This hypothesis was not supported. It would be interesting to investigate why investors reported that they have a higher satisfaction score with the recommendation system, intend to use it when it is commercially available but did not find it useful. The answers to such questions could go a long way in revealing more about the nature of collaborative filtering technology in the online investment domain.

Privacy is also an extremely important topic for future research. Privacy is closely related to the system security and trust discussed in this research, and is critical for the success of the recommendation system. Investors could potentially be concerned that the online brokerage firms could be tracking their investment behavior and use this data either as recommendations for other investors or for their own benefit. As this research area is very sensitive, it is up to firms to decide on their design decisions with regards to collaborative filtering and make investors comfortable with their decisions. Since collaborative filtering includes different approaches, each firm can custom tailor its recommendation engine in a very specific and unique way. This concern could be solved by ensuring investors that all recommendations would be anonymous while at the same anonymous while at the same time providing recommendations only to investors who agree on participating anonymously. Since collaborative filtering requires a database of pre-existing portfolios to generate recommendations, a trade-off must be agreed upon by investors and online brokerage firms as to how to manage and organize this relationship.

7. References

Albayrak, S., Wollny, S., Varone, N., Lommatzsch, A., & Milosevic, D. (2005). "Agent technology for personalized information filtering: The PIA system". Paper presented at the Santa Fe, New Mexico, USA,

Alpert, S. R., Karat, J., Karat, C., Brodie, C., & Vergo, J. G. (Nov 2003). "User attitudes regarding a user adaptive eCommerce web site". [Electronic version]. *User Modeling and User Adapted Interaction, 13*(4)

Ansari, A., Essegaier, S., & Kohli, R. (Aug 2000). "Internet recommendation systems". [Electronic version]. Journal of Marketing Research, 37(3)

Balasubramanian, S., Konana, P., & Menon, N. M. (July 2003). "Customer satisfaction in virtual environments: A study of online investing". *Management Science, 49*(7), 871.

Banerjee, S., Kauffman, R. J., & Wang, B. (2005). "A dynamic bayesian analysis of the drivers of internet firm survival". Paper presented at the Xi'an, China.

Barber, B. M., & Odean, T. (Winter 2001)." The Internet and the investor". [Electronic version]. *The Journal of Economic Perspectives, 15*(1), 41-54.

Barber, B. M., & Odean, T. (2002)."Online investors: Do the slow die first?" [Electronic version]. *The Review of Financial Studies, 15*(2 Special Issue), 455.

Barrett, M., & Scott, S. (2004). "Electronic trading and the process of globalization in traditional futures exchanges: A temporal perspective". [Electronic version]. *European Journal of Information Systems, 13*, 65.

Basilico, J., & Hofmann, T. (2004). "Unifying collaborative and content-based filtering." Paper presented at the Canada, Bellman, S., Lohse, G. L., & Johnson, E. J. (Dec 1999). "Predictors of online buying behavior". [Electronic version]. *Communications of the ACM, 42*(12), 32.

Cheung, K., & Tian, L. F. (Sep-Dec 2004). "Learning user similarity and rating style for collaborative recommendation". [Electronic version]. *Information Retrieval, 7*(3-4)

Choo, C. W., & Johnston, R. (2004)." Innovation in the knowing organization: A case study of E-commerce initiative". [Electronic version]. *Journal of Knowledge Management, 8*(6), 77.

Davis, F.D. (1989)."Perceived Usefulness, Ease of Use and User Acceptance of Information Technology". *MIS Quarterly*, Vol. 13 Issue 1 March 1989 pp. 319-339

Dellarocas, C. (2006). "The design of reliable trust management systems for electronic trading communities". *Unpublished manuscript.*

Delone, W.H and McLean E.R. (1992)."Information System Success: The Quest for Dependent Variable". *Information Systems Research,* Vol. 3, No.1 pp 60-95.

Dewally, M. (July/August 2003)."Internet investment advice: Investing with a rock of salt". [Electronic version]. *Financial Analysis Journal, 59*(4), 65.

Freund, C., & Weinhold, D. (May 2002)."The Internet and international trade in services." [Electronic version]. *The American Economic Review, 92*(2), 236.

Gallaugher, J., & Melville, N. (Aug 2004)."Electronic frontiers in foreign exchange trading". [Electronic version]. *Communications of the ACM, 47*(8), 81.

Golubski, W., Feuring, T., & Bernhard, J. (2000)." Web based money investment consultant ". Paper presented at the Crono, Italy, 797.

Hamers, L. & Y. Hemeryck & G. Herweyers (1989)."Similarity Measures in Scientometric Research: The Jaccard Index versus Salton's Cosine Formula". *Information Processing and Management.* 25(3).

Herlocker, J. Konstan, J. A., & Riedl, J. (October 2002)."An empirical analysis of design choices in neighborhood based collaborative filtering algorithms." [Electronic version]. *Information Retrieval, 5*(4)

Herlocker, J. L. Konstan, J. A., & Riedl, J. (2000)."Explaining collaborative filtering recommendations". Paper presented at the Philidelphia, PA, 24.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. (Jan 2004)."Evaluating collaborative filtering recommender systems". [Electronic version]. *ACM Transactions on Information Systems, 22*(1), 5.

Jin, R., Chai, J., & Si, L. (2004). "An automatic weighting scheme for collaborative filtering". Paper presented at the Sheffield, South Yorkshire, UK, 337.

Jin, R., & Si, L. (2004). "A study of methods for normalizing user ratings in collaborative filtering". Paper presented at the Sheffield, South Yorkshire, UK,

Jin, X., Zhou, Y., & Mobasher, B. (2005). "A maximum entropy web recommendation system: Combining collaborative and content features". Paper presented at the Chicago, Illinois, USA, 612.

Jin, X., Zhou, Y., & Mobasher, B. "Task oriented user modeling for recommendations". [Electronic version]. *Center for Web Intelligence, School of Computer Science, DePaul University USA,*

Jordan, A. (Dec 2001). "Five critical practices in mastering E-business". [Electronic version]. *Consulting to Management, 12*(4), 24.

Kavassalis, P., Lelis, S., Rafea, M., & Haridi, S. (Feb 2004)." What makes a web site popular?" [Electronic version]. *Communications of the ACM, 47*(2), 51.

Kohrs, A., & Merialdo, B. (1999)." Improving collaborative filtering with multimedia indexing techniques to create user-adaptive web sites". [Electronic version]. *ACM Multimedia, Institut EURECOM,* , 27.

Konana, P., Menon, N. M., & Balasubramanian, S. (January 2000)."The implications of online investing". [Electronic version]. *Communications of the ACM, 43*(1), 35.

Konstan, J. A. (Jan 2004)."Introduction to recommender systems: Algorithms and evaluation". [Electronic version]. *ACM Transactions on Information Systems, 22*(1),1.

Levinsohn, A. (Sep 1999). "Online brokerage: The new core account?" [Electronic version]. *American Bankers Association ABA Banking Journal, 91*(9)

Lin, F., Huang, S., & Chen, N. (Winter 2004). "Incremental revision of recommendation rules for information services". [Electronic version]. *E-Service Journal, 3*(2), 85.

Maxemchuk, N. F., & Shur, D. H. (August 2001). "An internet multicast system for the stock market". [Electronic version]. *ACM Transactions on Computer Systems, 19*(3), 384.

McCarthy, K., Salamo, M., Coyle, L., McGinty, L., Smyth, B., & Nixon, P. (2006). "Group recommender systems: A critiquing based approach". Paper presented at the Sydney, Australia,

Mild, A., & Natter, M. (June 2002). "Collaborative filtering or regression models for internet recommendation systems?" [Electronic version]. *Journal of Targeting, Measurement and Analysis for Marketing, 10*(4), 304.

Miller, B. N., Konstan, J., A., & Riedl, J. (July 2004)." PocketLens: Toward a personal recommender system". [Electronic version]. *ACM Transactions on Information Systems, 22*(3), 437.

Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2002)." Discovery and evaluation of aggregate usage profiles for web personalization". [Electronic version]. *Data Mining and Knowledge Discovery, 6*, 61.

Neumann, P. G., & Weinstein, L. (Nov 1999). "Risks of content filtering" [Electronic version]. *Association for Computing Machinery, Communications of the ACM, 42*(11), 152.

Nguyen, V. T., Van Ness, Bonnie F., & Van Ness, Robert A. (Fall 2004)." The reporting of island trades on the Cincinnati stock exchange". [Electronic version]. *Journal of Applied Finance, 14*(2), 30.

O'Donnovan, J., & Smyth, B. (2006)."Is trust robust? An analysis of trust based recommendation". Paper presented at the Sydney, Australia, 101.

O'Mahony, M. P., Hurley, N. J., & Silvestre, G. C. M. (2006). "Detecting noise in recommender system databases" Paper presented at the Sydney, Australia, 109.

Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2005). *Introduction to Data Mining* (2005). Addison Wesley; US Ed edition.

Peterffy, T., & Battan, D. M. (Aug 2004). "Why some dealers and exchanges have been slow to automate". [Electronic version]. *Financial Analysis Journal, 60*(4), 15.

Pu, P., & Chen, L. (2006). "Trust building with explanation interfaces." Paper presented at the Sydney, Australia,

Rieley, Frank K., Keith C. Brown. *Investment Analysis and Portfolio Management.* USA: South Western, a division of Thomson Learning, 2003 (7th Edition)

Saatcioglu, K., Stallaert, J., & Whinston, A. B. (June 2001). "Design of a financial portal."[Electronic version]. *Communications of the ACM, 44*(6)

Srikumar, K. (Apr-Jun 2004)." A framework of agent-based personalized recommender system for E-commerce". [Electronic version]. *South Asian Journal of Management, 11*(2), 66.

Stallaert, M. F., & Whinston, A. B. (Nov 2000). "The internet and the future of financial markets". [Electronic version]. *Communications of the ACM, 43*(11), 83.

Strader, T. J., & Ramaswami, S. N. (July 2004). "Investor perceptions of traditional and online channels". [Electronic version]. *Communications of the ACM, 47*(7)

Tseng, C. (2004)." Portfolio management using hybrid recommendation system." Paper presented at the

Vucetic, S., & Obradovic, Z. (2005). "Collaborative filtering using a regression based approach". [Electronic version]. *Knowledge and Information Systems, 7*, 1.

Wade, Michael R & Peter Tingling. (Summer 2005). "A New Twist on an Old Method: A Guide to the Applicability and Use of Web Experiments in Information System Research." *Databases for Advances in Information Systems,* 36(3).

WebSM- Web Survey Methodology. 15 September 2006. **http://www.websm.org/uploadi/editor/1133720534GUIDE_1_Frames_and_sampling_in_web_surveys.pdf**

Weng, L., Xu, Y., & Li, Y. (2005)." A framework for E-commerce oriented recommendation systems." Paper presented at the 309.

Wikipedia The Free Encyclopedia. 14 April 2006. **http://www.wikipedia.com**

World Internet Usage Statistics. 15 May 2006. **http://www.internetworldstats.com/stats.htm**

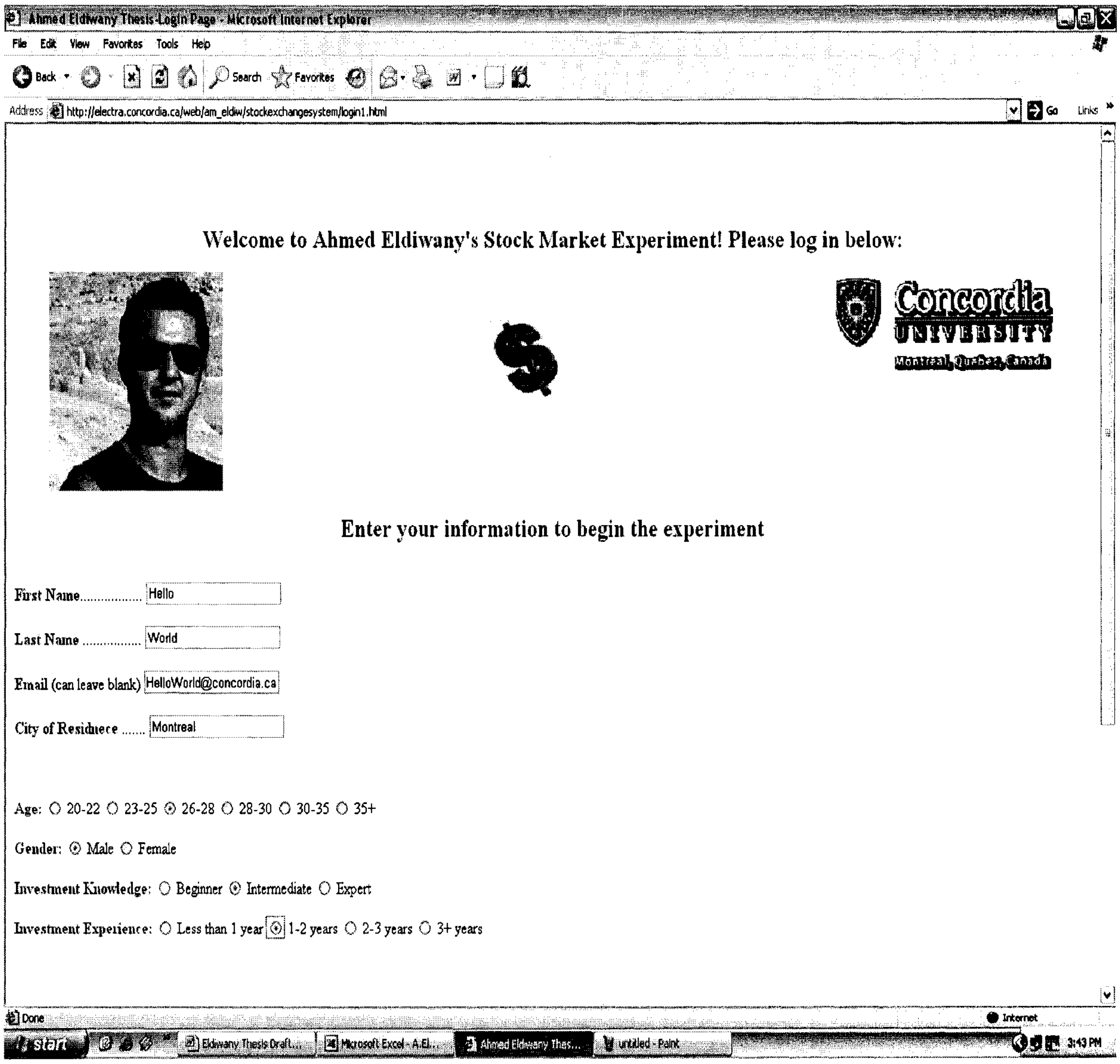
Xue, G., Lin, C., Yang, Q., Xi, W., Zeng, H., & Yu, Y. et al. (2005). "Scalable collaborative filtering using cluster based smoothing. "Paper presented at the Salvador, Brazil, 114.

Zitnick, C. L., & Kanade, T. (2004). "Maximum entropy for collaborative filtering." Paper presented at the 636 Conferece.

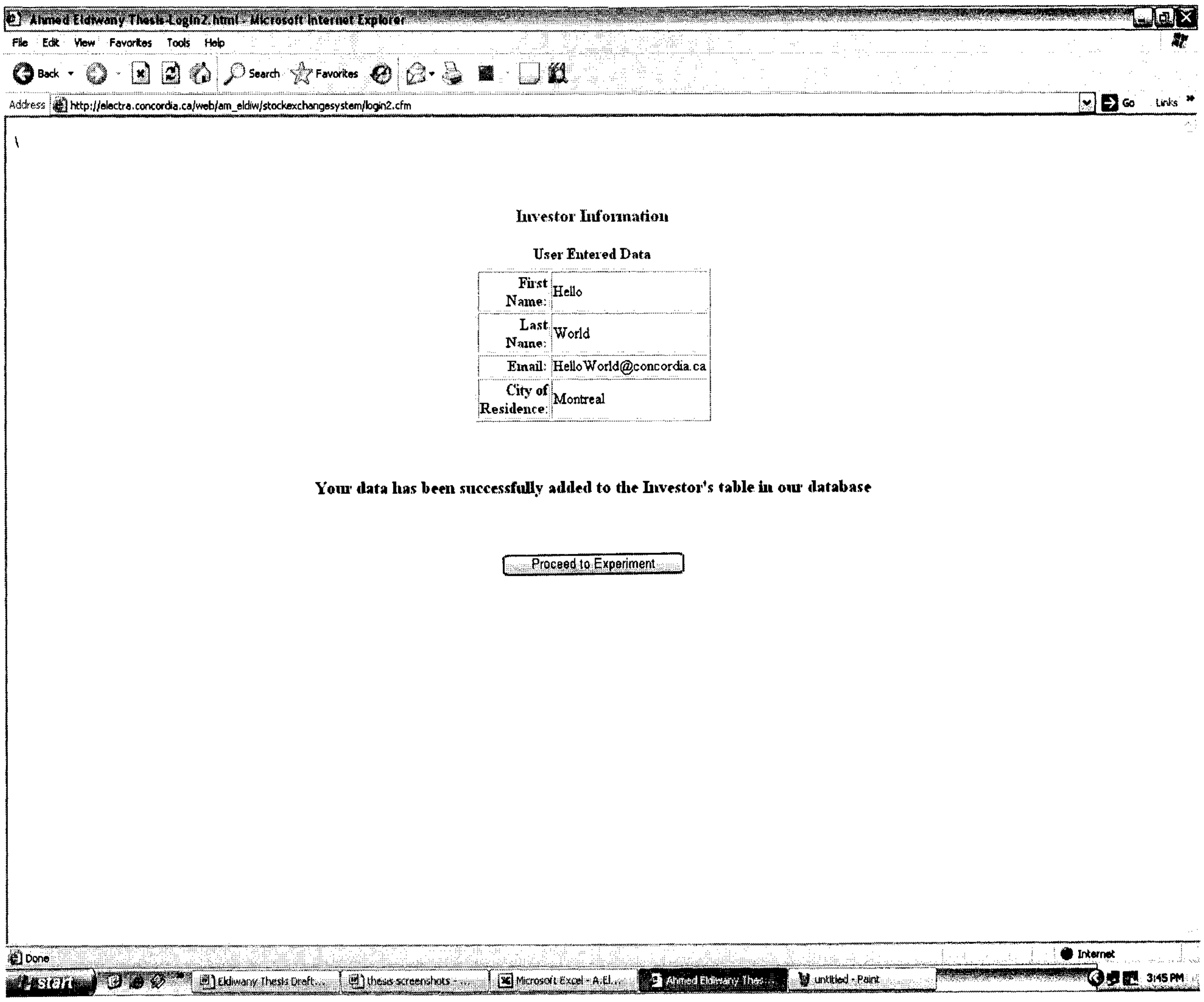
8. Appendix

8.1 System Screenshots

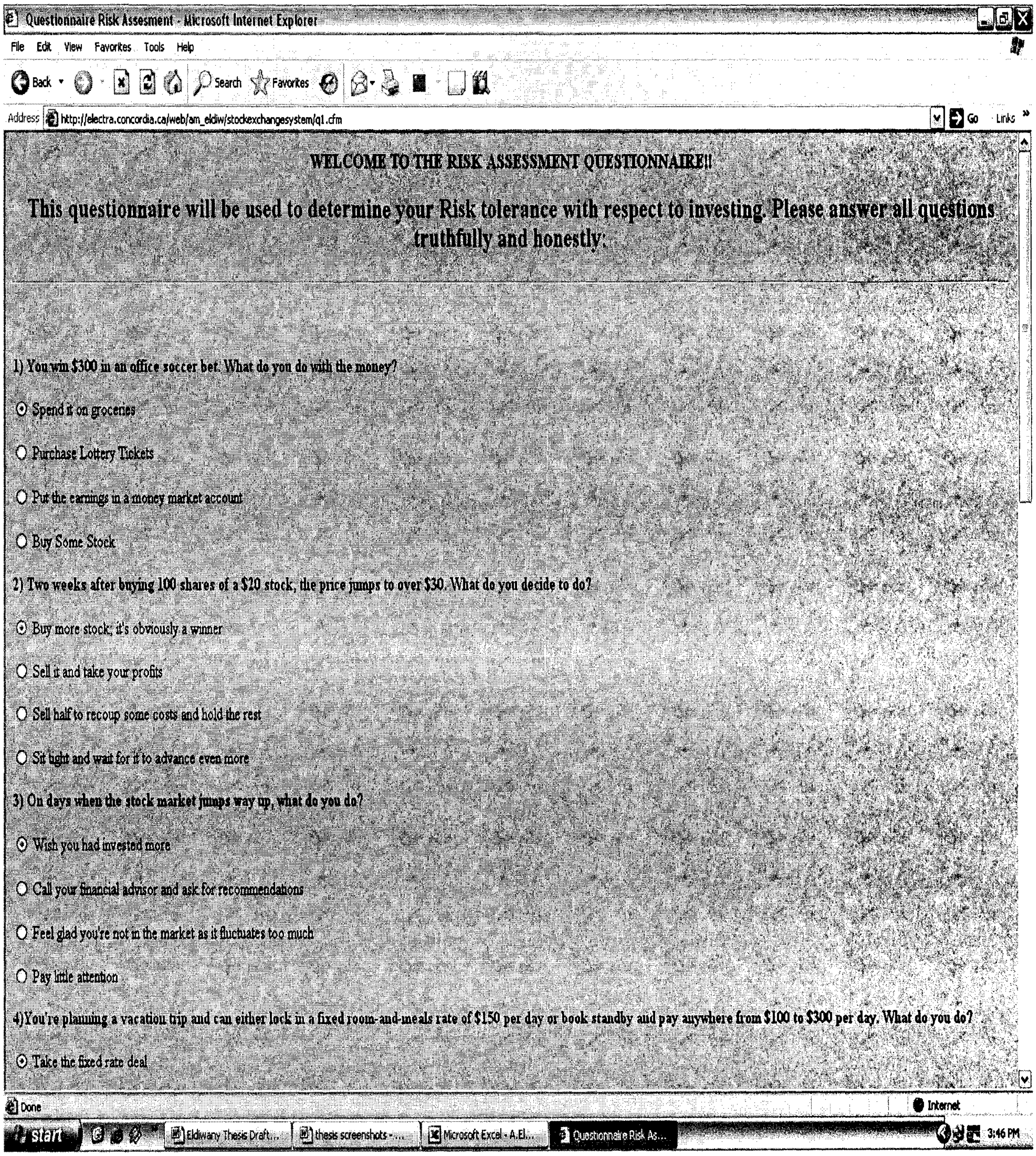
8.1.1-Login Page-



8.1.2 System Screenshots-Data Validation Page-



8.1.3 Risk Assessment introduction and Questionnaire


Questionnaire Risk Assesment - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://electra.concordia.ca/web/am_eldiw/stockexchangesystem/q1.cfm
Go Links
WELCOME TO THE RISK ASSESSMENT QUESTIONNAIRE!!
This questionnaire will be used to determine your Risk tolerance with respect to investing. Please answer all questions truthfully and honestly.
1) You win $300 in an office soccer bet. What do you do with the money?
Spend it on groceries
Purchase Lottery Tickets
Put the earnings in a money market account
Buy Some Stock
2) Two weeks after buying 100 shares of a $20 stock, the price jumps to over $30. What do you decide to do?
Buy more stock; it's obviously a winner
Sell it and take your profits
Sell half to recoup some costs and hold the rest
Sit tight and wait for it to advance even more
3) On days when the stock market jumps way up, what do you do?
Wish you had invested more
Call your financial advisor and ask for recommendations
Feel glad you're not in the market as it fluctuates too much
Pay little attention
4)You're planning a vacation trip and can either lock in a fixed room-and-meals rate of $150 per day or book standby and pay anywhere from $100 to $300 per day. What do you do?
Take the fixed rate deal
Done
Internet
start
Eldiwany Thesis Draft...
thesis screenshots - ...
Microsoft Excel - A.El...
Questionnaire Risk As...
3:46 PM

<u>**Risk Assessment Questionnaire (actual questions)**</u>

(Rieley and Brown, 2003)

1) You win $300 in an office soccer bet. What do you do with the money?

a) ☐ Spend it on groceries
b) ☐ Purchase Lottery Tickets
c) ☒ Put the earnings in a money market account
d) ☐ Buy Some Stock

2) Two weeks after buying 100 shares of a $20 stock, the price jumps to over $30. What do you decide to do?

a) ☒ Buy more stock; it's obviously a winner
b) ☐ Sell it and take your profits
c) ☐ Sell half to recoup some costs and hold the rest
d) ☐ Sit tight and wait for it to advance even more

3) On days when the stock market jumps way up, what do you do?

a) ☒ Wish you had invested more
b) ☐ Call your financial advisor and ask for recommendations
c) ☐ Feel glad you're not in the market as it fluctuates too much
d) ☐ Pay little attention

4)You're planning a vacation trip and can either lock in a fixed room-and-meals rate of $150 per day or book standby and pay anywhere from $100 to $300 per day. What do you do?

a) ☐ Take the fixed rate deal
b) ☐ Talk to people who have been there about the availability of last minute accommodations
c) ☐ Book standby and also arrange vacation insurances because you're tired of the tour operator
d) ☒ Take your chances with standby

5) The owner of your apartment building is converting the units to condominiums. You can buy your unit for $75,000 or an option on a unit for $15,000. (Units have recently sold for close to $100,000 and prices seem to be going up). For financing you'll have to borrow the down payment and pay mortgage and condo fees higher than your present rent. What do you do

a) ☐ Buy your unit
b) ☐ Buy your unit and look for another unit to buy
c) ☒ Sell your option and arrange to rent the unit yourself
d) ☐ Sell the option and move out because you think the conversion will attract couples with small children (and lots of noise!)

6) You have been working for 3 years for a rapidly growing company. As an executive, you are offered the option of buying up to 2% of company stock: 2,000 shares at $10 a share. Although the company is privately owned (its stock does not trade in the open market), it's majority owner has made handsome profits selling three other businesses and intends to sell this one eventually. What do you do?

a) ☐ Purchase all the shares you can and tell the owner you would invest more if allowed
b) ☐ Purchase all the shares
c) ☒ Purchase half the shares
d) ☐ Purchase a small amount of shares

7) You go to a casino for the first time. You chose to play which game?

a) ☒ Quarter slot machines
b) ☐ $5 minimum bet roulette
c) ☐ S$1 slot machine
d) ☐ $25 minimum bet blackjack

8) You want to take someone out for a special dinner in a city that's new to you. How do you pick a place?

a) ☐ Read restaurant reviews in the local newspaper
b) ☐ Ask coworkers if they now of a suitable place
c) ☐ Call the only other person you know in this city, who eats out a lot but only recently moved here

d) ☐ Visit the city sometime before your dinner to check out the restaurants yourself.

9) The expression that best describes your lifestyle is:

a) ☐ No guts, no glory
b) ☐ Just do it!
c) ☐ Look before you leap
d) ☐ All good things come to those who wait

10) Your attitude toward money is best described as

a) ☐ A dollar saved is a dollar earned
b) ☐ You've got to spend money to make money
c) ☐ Cash and carry only
d) ☐ Whenever possible, use other people's money

The scoring system for this questionnaire is based on the summation of the points for each answer. This questionnaire is used by practitioners in the brokerage industry nowadays and provides a good assessment of the investor's risk attitude with regards to investment and classifies the investor into one of the following four domains:

1. Conservative (total points of 10-17)
2. Semi Conservative (total points of 18-25)
3. Semi Aggressive (total points of 26-32)
4. Aggressive (33-40)

The breakdown of the points is as follows for each question:

1) a-1, b-4, c-2, d-3

2) a-4, b-1, c-3, d-2

3) a-3, b-4, c-2, d-1

4) a-2, b-3, c-1, d-4

5) a-3, b-4, c-2, d-1

6) a-4, b-3, c-2, d-1

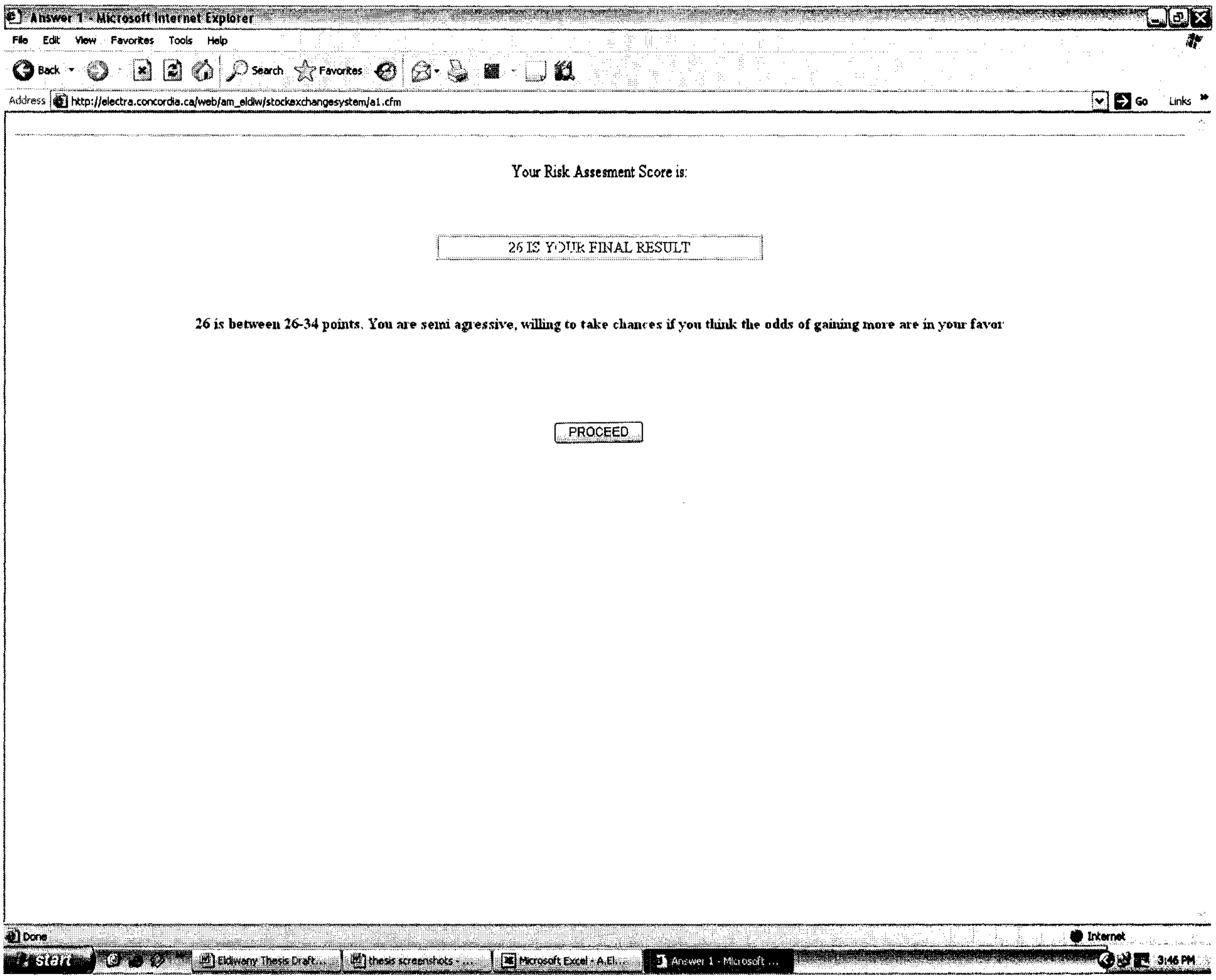
7) a-1, b-3, c-2, d-4

8) a-2, b-3, c-4, d-1

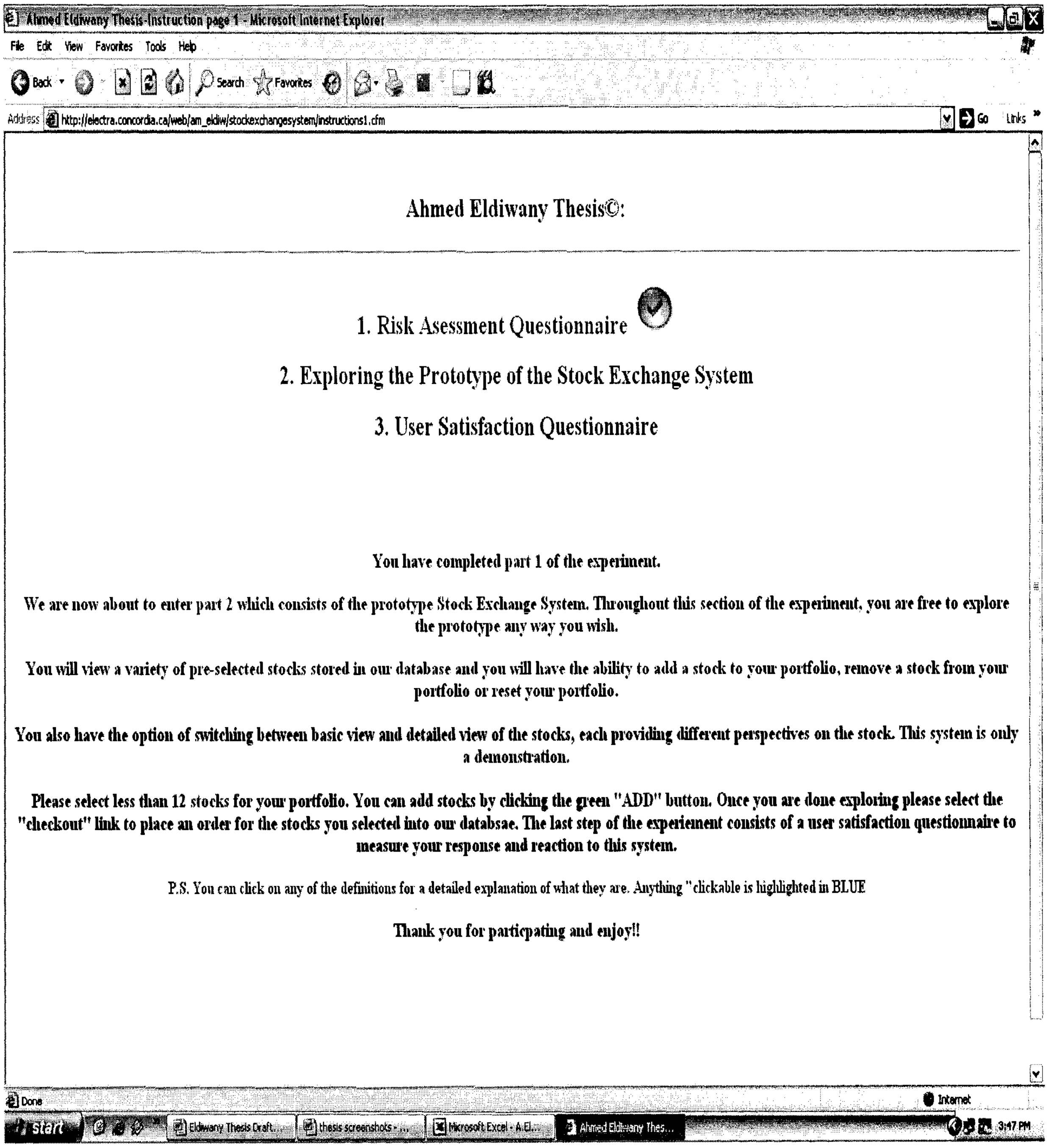
9) a-4, b-3, c-2, d-4

10) a-2, b-3, c-1, d-4

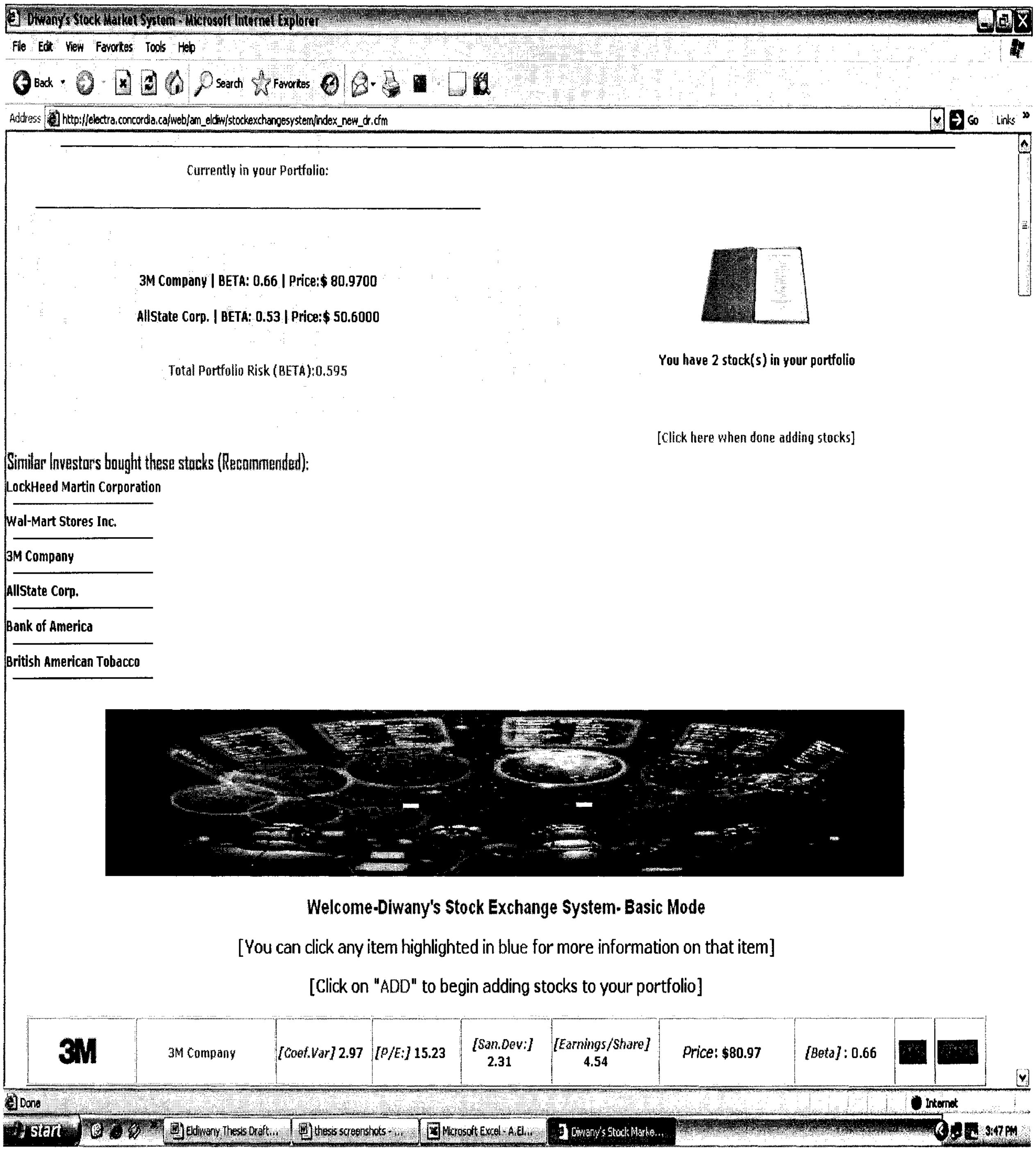
8.1.4 System Screenshots-Risk Assessment Score-

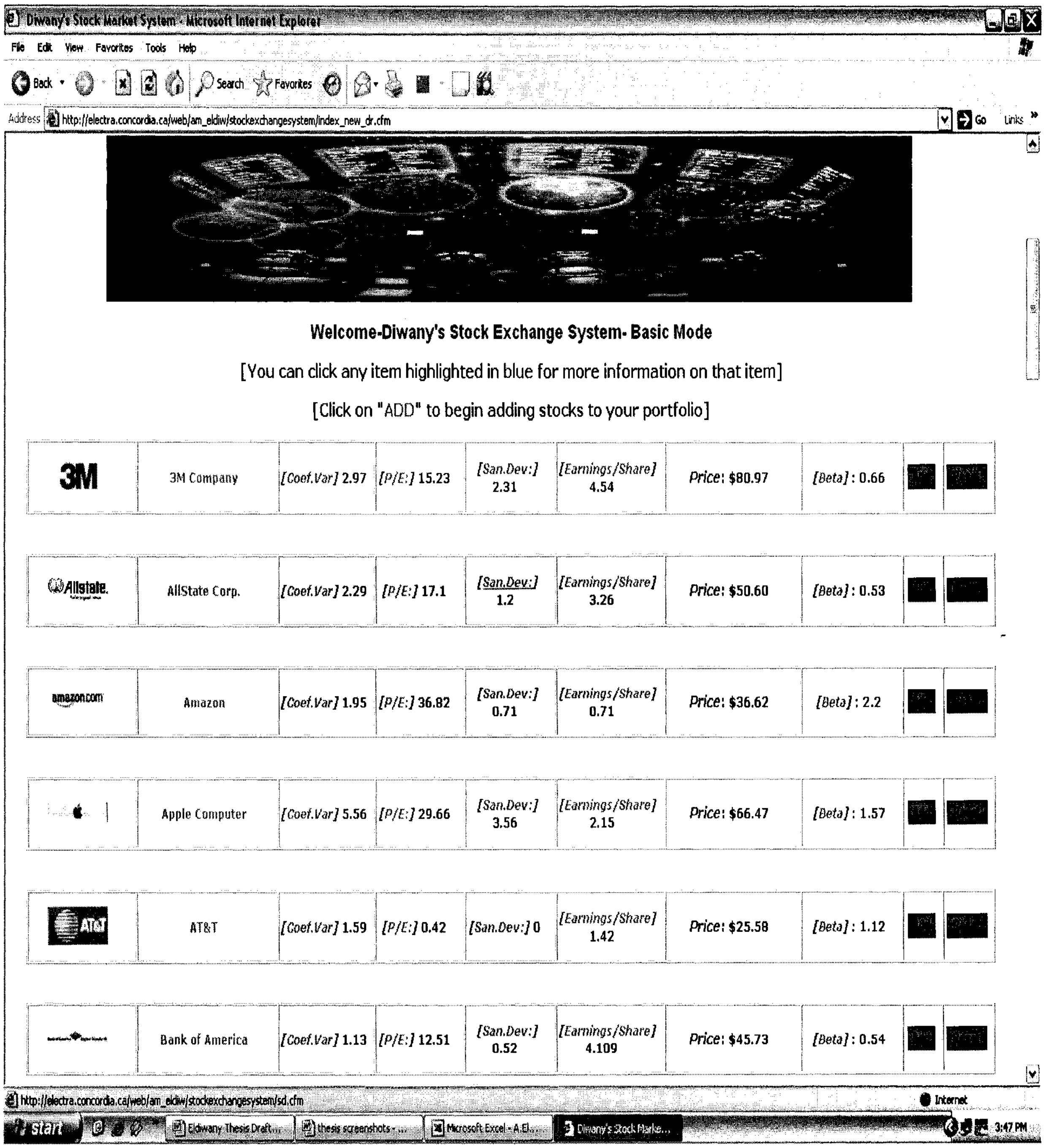


8.1.5 System Screenshots-Progress Report Page-



8.1.6 System Screenshots-Main System Sample Page-





Diwany's Stock Market System - Microsoft Internet Explorer

Address http://electra.concordia.ca/web/am_eldiw/stockexchangesystem/index_new_dr.cfm

Welcome-Diwany's Stock Exchange System- Basic Mode

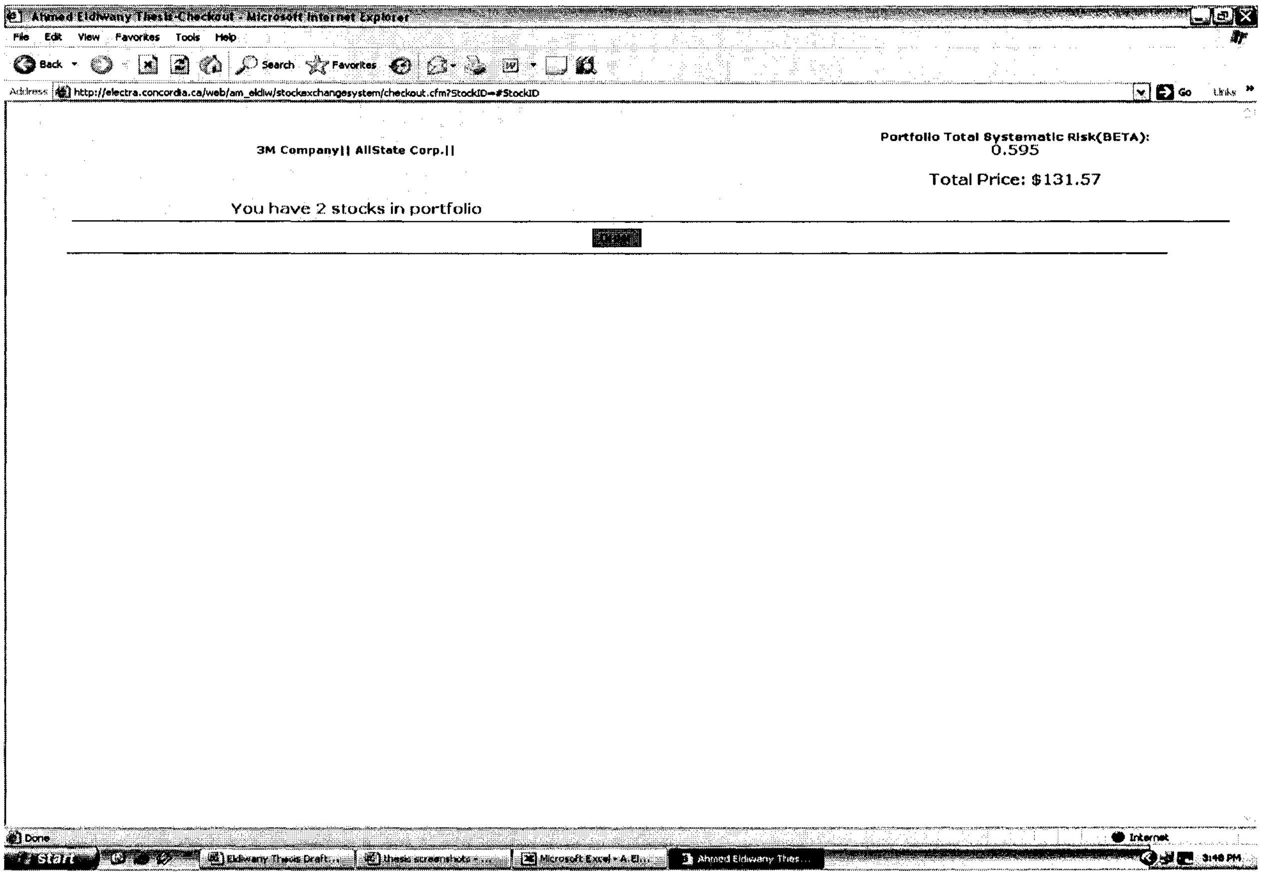
[You can click any item highlighted in blue for more information on that item]

[Click on "ADD" to begin adding stocks to your portfolio]

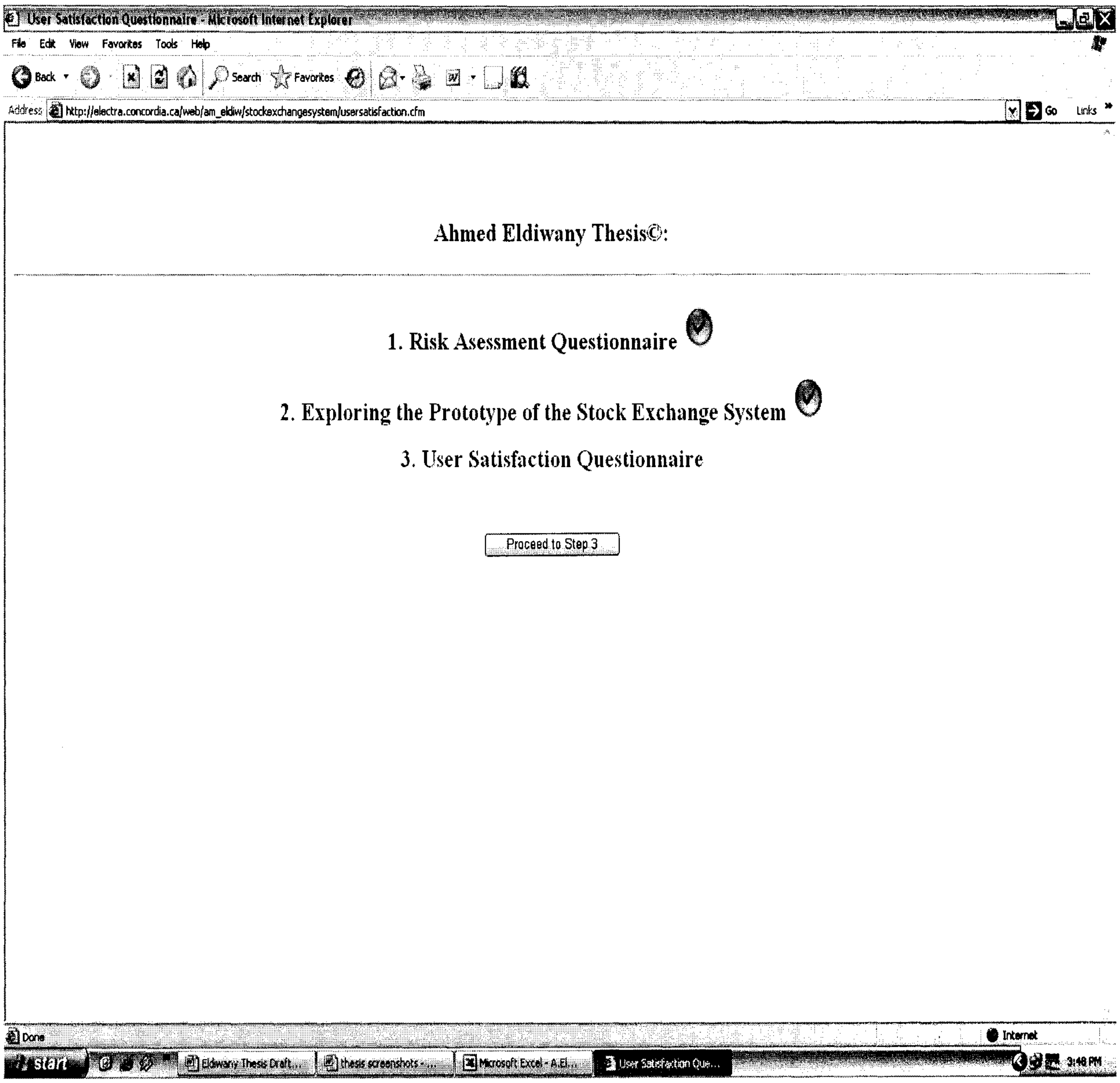
3M	3M Company	[Coef.Var] 2.97	[P/E:] 15.23	[San.Dev:] 2.31	[Earnings/Share] 4.54	Price: $80.97	[Beta]: 0.66
Allstate	AllState Corp.	[Coef.Var] 2.29	[P/E:] 17.1	[San.Dev:] 1.2	[Earnings/Share] 3.26	Price: $50.60	[Beta]: 0.53
amazon.com	Amazon	[Coef.Var] 1.95	[P/E:] 36.82	[San.Dev:] 0.71	[Earnings/Share] 0.71	Price: $36.62	[Beta]: 2.2
	Apple Computer	[Coef.Var] 5.56	[P/E:] 29.66	[San.Dev:] 3.56	[Earnings/Share] 2.15	Price: $66.47	[Beta]: 1.57
AT&T	AT&T	[Coef.Var] 1.59	[P/E:] 0.42	[San.Dev:] 0	[Earnings/Share] 1.42	Price: $25.58	[Beta]: 1.12
	Bank of America	[Coef.Var] 1.13	[P/E:] 12.51	[San.Dev:] 0.52	[Earnings/Share] 4.109	Price: $45.73	[Beta]: 0.54

http://electra.concordia.ca/web/am_eldiw/stockexchangesystem/sd.cfm

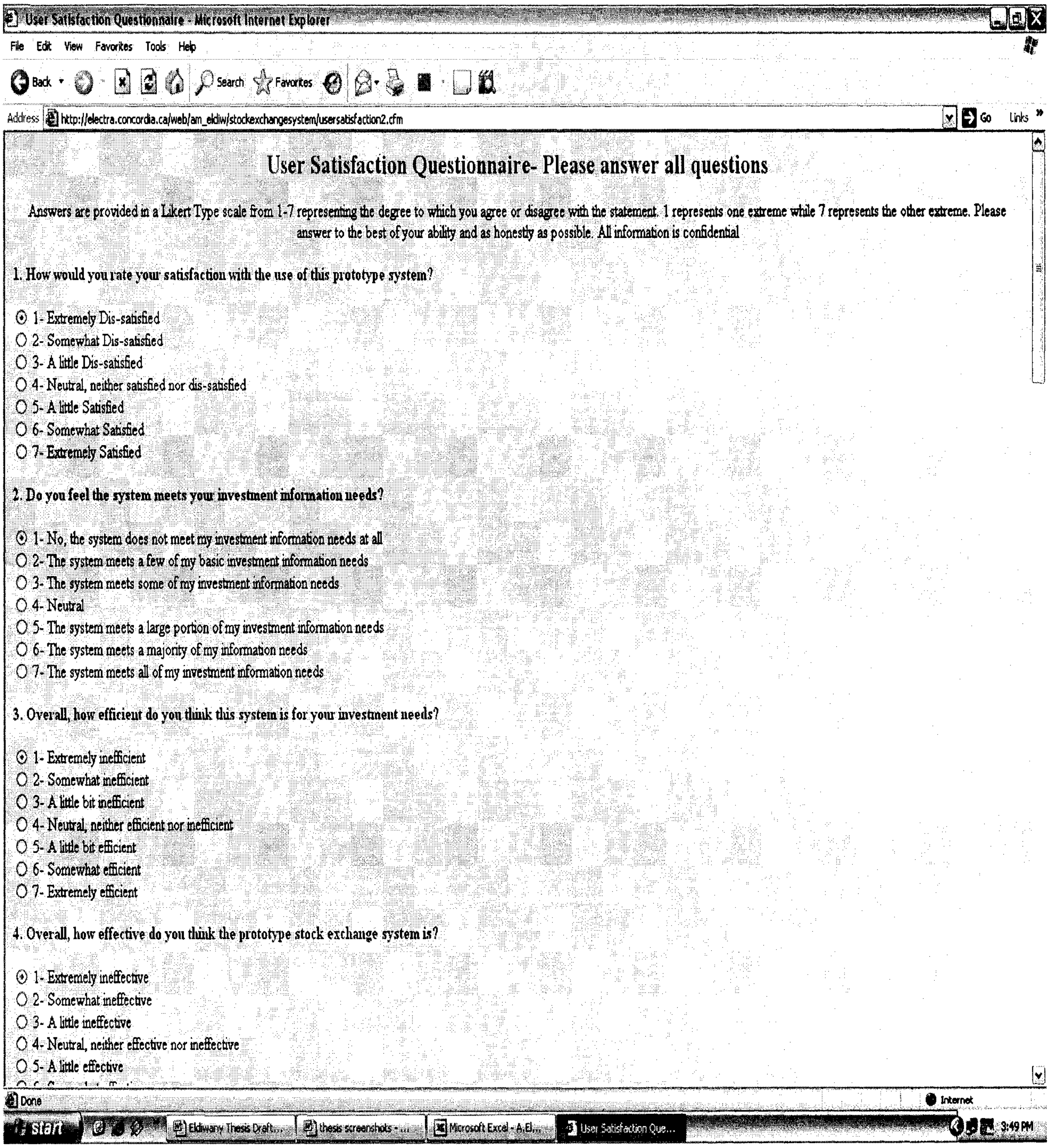
8.1.7 System Screenshots-Portfolio Confirmation-



8.1.8 System Screenshots- Second Progress Report Page-



8.1.9 User Satisfaction post-test questionnaire:


User Satisfaction Questionnaire - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://electra.concordia.ca/web/am_eldiw/stockexchangesystem/usersatisfaction2.cfm

User Satisfaction Questionnaire- Please answer all questions

Answers are provided in a Likert Type scale from 1-7 representing the degree to which you agree or disagree with the statement. 1 represents one extreme while 7 represents the other extreme. Please answer to the best of your ability and as honestly as possible. All information is confidential

1. How would you rate your satisfaction with the use of this prototype system?

- 1- Extremely Dis-satisfied
- 2- Somewhat Dis-satisfied
- 3- A little Dis-satisfied
- 4- Neutral, neither satisfied nor dis-satisfied
- 5- A little Satisfied
- 6- Somewhat Satisfied
- 7- Extremely Satisfied

2. Do you feel the system meets your investment information needs?

- 1- No, the system does not meet my investment information needs at all
- 2- The system meets a few of my basic investment information needs
- 3- The system meets some of my investment information needs
- 4- Neutral
- 5- The system meets a large portion of my investment information needs
- 6- The system meets a majority of my information needs
- 7- The system meets all of my investment information needs

3. Overall, how efficient do you think this system is for your investment needs?

- 1- Extremely inefficient
- 2- Somewhat inefficient
- 3- A little bit inefficient
- 4- Neutral, neither efficient nor inefficient
- 5- A little bit efficient
- 6- Somewhat efficient
- 7- Extremely efficient

4. Overall, how effective do you think the prototype stock exchange system is?

- 1- Extremely ineffective
- 2- Somewhat ineffective
- 3- A little ineffective
- 4- Neutral, neither effective nor ineffective
- 5- A little effective

Done | Internet

start | Eldiwany Thesis Draft... | thesis screenshots - ... | Microsoft Excel - A.El... | User Satisfaction Que... | 3:49 PM

User Satisfaction Questionnaire (actual questions)

1. **How would you rate your satisfaction with the use of the system?**

   1 2 3 4 5 6 7

2. **Do you feel the system meets your investment information needs?**

   1 2 3 4 5 6 7

3. **Overall, how efficient do you think this system is?**

   1 2 3 4 5 6 7

4. **Overall, how effective do you think the stock exchange system is?**

   1 2 3 4 5 6 7

5. **Overall, how satisfied are you with using this system as it is?**

   1 2 3 4 5 6 7

6. **Using this system, my investment tasks would be accomplished more quickly.**

   1 2 3 4 5 6 7

7. **Using this stock exchange system would improve my investment performance**

   1 2 3 4 5 6 7

8. **Using this system would make investment easier for me**

   1 2 3 4 5 6 7

9. **I would find this stock exchange system useful in my investment decisions**

   1 2 3 4 5 6 7

10. **Learning to work with this system would be easy for me**

    1 2 3 4 5 6 7

11. I would find it easy to get this system to do what I want it to do

1 2 3 4 5 6 7

12. It would be easy for me to become skillful at using this system

1 2 3 4 5 6 7

13. I would find this stock exchange system easy to use

1 2 3 4 5 6 7

14. I plan on using this stock exchange system once it is released

1 2 3 4 5 6 7

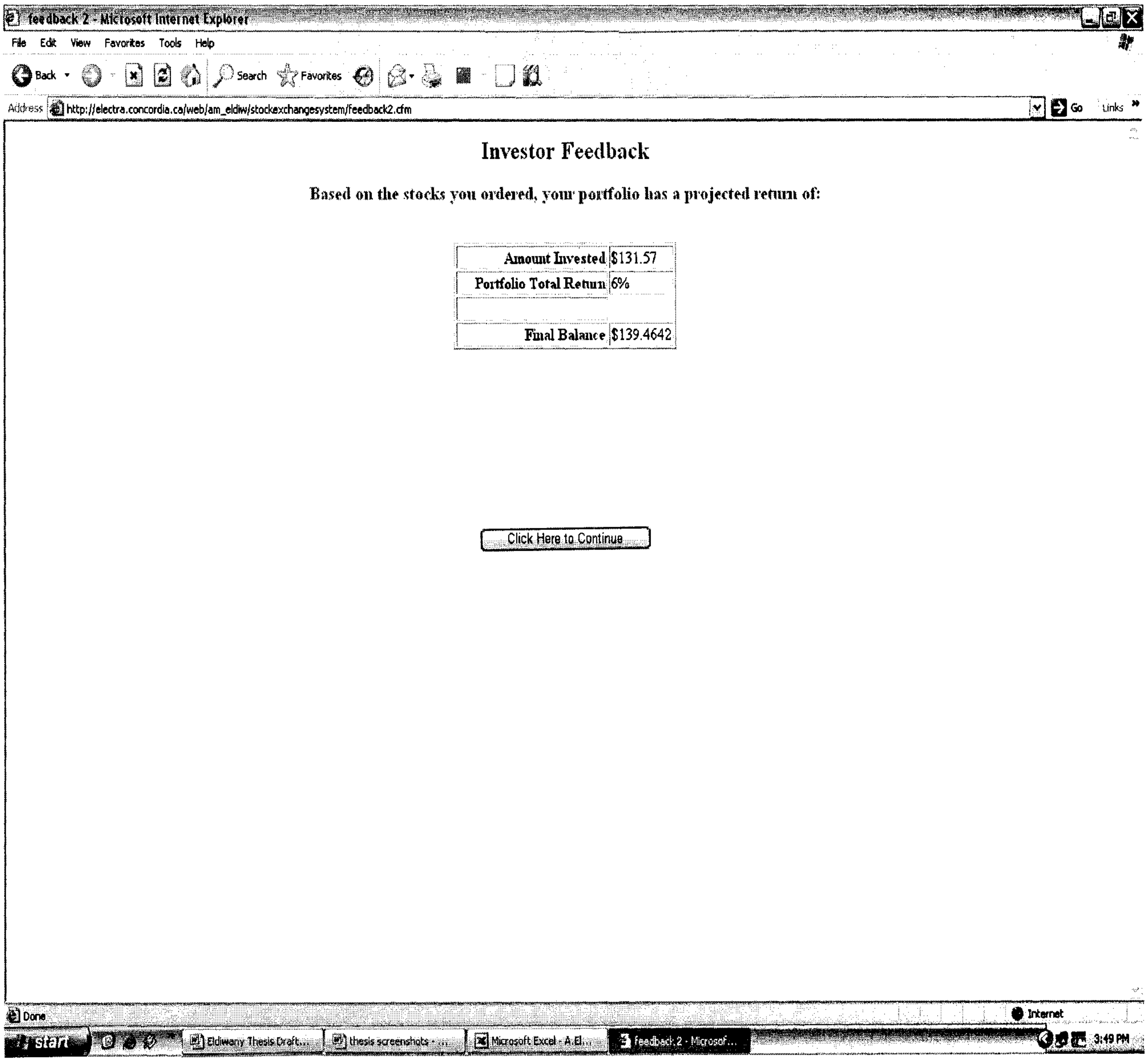
15. I intend to use this system frequently when it is available

1 2 3 4 5 6 7

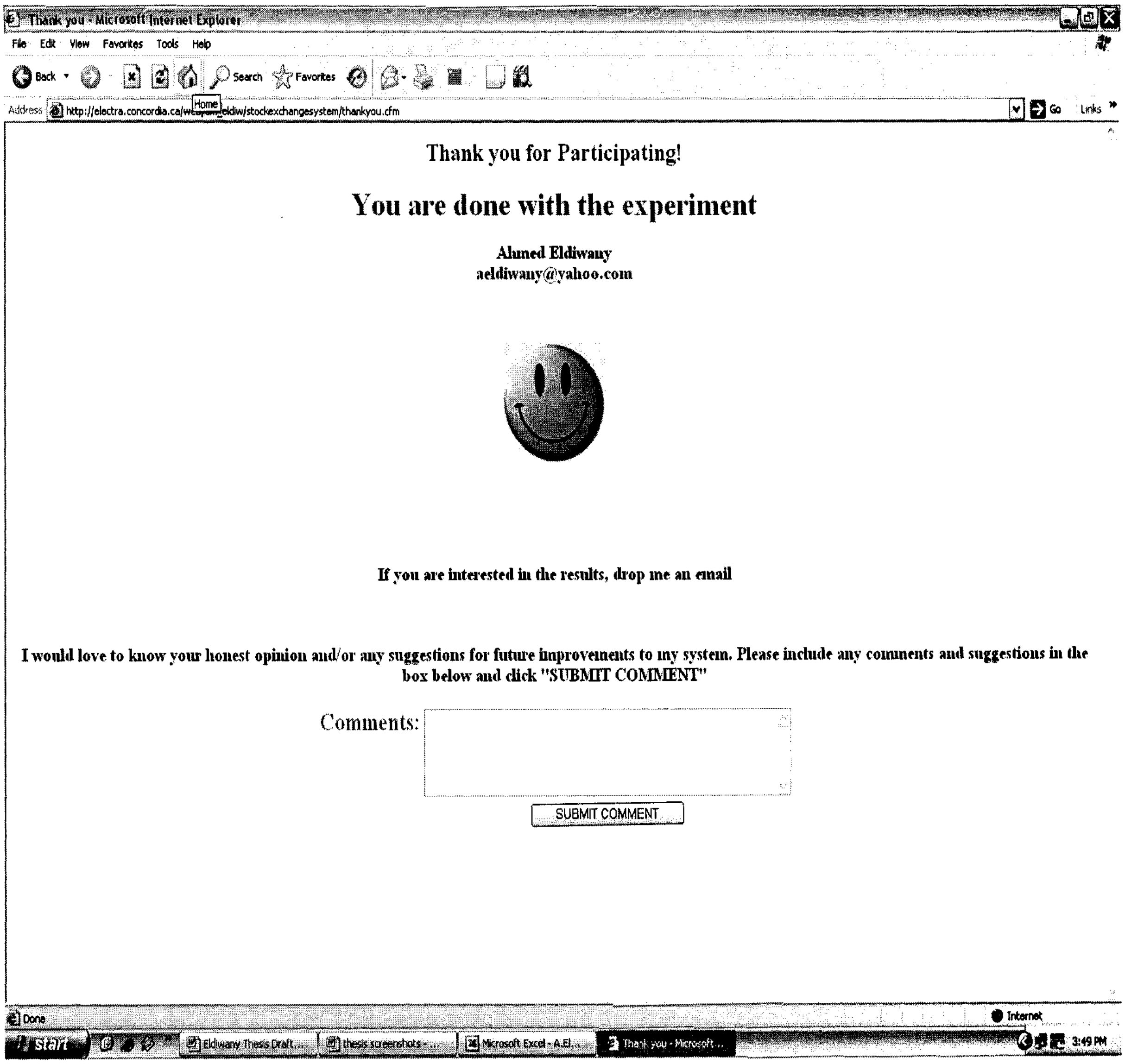
16. I intend to be a heavy user of this system once it is online

1 2 3 4 5 6 7

8.1.10 System Screenshots-Investor Feedback-



8.1.11 System Screenshots- Thank You Page-



8.2 SPSS Output

8.2.1 Rotated Component Matrices

Rotated Component Matrix [a]

	Component		
	1	2	3
Q1	.828	.077	.288
Q2	.673	-.007	.616
Q3	.834	.274	.219
Q4	.836	.047	-.121
Q5	.279	.728	.274
Q6	.163	.919	-.022
Q7	.779	.194	.082
Q8	.150	.787	.376
Q9	.748	.490	-.069
Q10	.031	.612	.559
Q11	.711	.169	.462
Q12	.092	.355	.832

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Rotated Component Matrix[a]

	Component		
	1	2	3
Q1	.846	.078	.286
Q2	.701	-.027	.583
Q3	.843	.297	.207
Q4	.882	.080	-.195
Q6	.148	.941	.059
Q8	.134	.776	.441
Q10	.059	.569	.646
Q12	.130	.239	.846

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

8.2.2 Reliability Statistics

1. Satisfaction Cronbach's Alpha

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.877	.876	4

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.973	4.061	5.394	1.333	1.328	.388	4

The covariance matrix is calculated and used in the analysis.

2. Usefulness Cronbach's Alpha

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.826	.827	2

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.864	5.712	6.015	.303	1.053	.046	2

The covariance matrix is calculated and used in the analysis.

3. Intention to Use Cronbach's Alpha

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.694	.735	2

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.871	4.424	5.318	.894	1.202	.400	2

The covariance matrix is calculated and used in the analysis.

8.2.3 Independent T-Tests

1. **Satisfaction**

Group Statistics

	System	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction	0	29	4.4655	1.43571	.26660
	1	37	5.3716	1.21424	.19962

Independent Samples Test

		Levene's Test for quality of Variance		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	ig. (2-tailed	Mean Difference	Std. Error Difference	Lower	Upper
Satisfacti	Equal varian assumed	.762	.386	-2.777	64	.007	-.90610	.32632	1.55800	-.25421
	Equal varian not assumed			-2.721	54.799	.009	-.90610	.33306	1.57362	-.23859

2. Usefulness

Group Statistics

	System	N	Mean	Std. Deviation	Std. Error Mean
Usefulness	0	29	5.7586	1.10696	.20556
	1	37	5.9459	1.28983	.21205

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Usefulness	Equal variances assumed	.406	.526	-.623	64	.536	-.18733	.30089	-.78843	.41378
	Equal variances not assumed			-.634	63.431	.528	-.18733	.29533	-.77741	.40276

3. Intention to Use

Group Statistics

	System	N	Mean	Std. Deviation	Std. Error Mean
ITU	0	29	4.4655	1.37536	.25540
	1	37	5.1892	1.17468	.19312

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ITU	Equal variances assumed	.117	.734	-2.304	64	.024	-.72367	.31408	1.35112	-.09622
	Equal variances not assumed			-2.260	55.149	.028	-.72367	.32019	1.36531	-.08203

8.2.4 Measuring statistical differences

1-Age t tests

Group Statistics

	System	N	Mean	Std. Deviation	Std. Error Mean
Age	0	29	3.2414	1.32706	.24643
	1	37	3.3514	1.43790	.23639

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Age	Equal variances assumed	.423	.518	-.319	64	.751	-.10997	.34486	-.79891	.57896
	Equal variances not assumed			-.322	62.246	.748	-.10997	.34148	-.79252	.57258

2-Investment Experience

Group Statistics

	System	N	Mean	Std. Deviation	Std. Error Mean
XP	0	29	1.9655	1.11748	.20751
	1	37	1.9189	1.03758	.17058

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
XP	Equal variances assumed	.362	.549	.175	64	.862	.04660	.26618	-.48516	.57836
	Equal variances not assumed			.173	58.020	.863	.04660	.26862	-.49110	.58430

3-Investment Knowledge

Group Statistics

	System	N	Mean	Std. Deviation	Std. Error Mean
K	0	29	1.5862	.73277	.13607
	1	37	1.5405	.60528	.09951

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
K	Equal variances assumed	1.867	.177	.277	64	.782	.04567	.16470	-.28336	.37469
	Equal variances not assumed			.271	53.954	.788	.04567	.16857	-.29231	.38364