ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 3, May 2012

E-Investment Banking: NextGen Investment

Prof. D. R. Ingle, Dr. B. B. Meshram

Abstract— **Investment is the biggest risky challenge in the current market. World Wide Web has experienced remarkable growth in recent years hence security has become one of the major promising task in the present scenario of e-business environment. Investment is an asset or item that is utilized with a hope to generate income in the future. Therefore investment is the basic need for everyone to secure future. Web investor can invest their capital in various sectors however the success would depend upon the appropriate choice based on the actual knowledge input about the investment sector. This paper emphasizes the best choice for investors to invest their capitals using web services for communication purpose and Data mining concept to extract knowledge from a data set in a human-understandable structure from its previous history and also protection mechanisms for the investors. We are using data mining algorithm Decision Tree and C5 algorithm for comparison and making decision. Using web service we can get online stock / share market data (using live stock quotes) and can display on my web page. History DB will contains 5 years history information of shares and then it compared with History DB by using data mining algorithm C5 algorithm and make prediction by using**
Decision Tree & update the same and to maintain the security we are using SHA-1 algorithm.

***Index Terms*— Cash Prediction, Investment, Banking, Web Services,**

I. INTRODUCTION

People who know many leaps and bounds of the investment market due to their perceiving experience in the market but fail to make most of it due to lack of capability to invest. There is very limited possible ways where these two people will meet and achieve each other's trust. Imagine the situation when, in recession, people relied solely on their job and when the market went down they had absolutely nothing, and then realized that there is problem. Financial magazines stated that it would happen often because of fluctuating market policies. Banking websites on the other side can reduce the chance of a dead investment but they also tend to consume most of the profits. Social networking, financial blogs to some extent helped but investees found it tedious because of lack of time and idea security.[1] The major focus of the application is on bringing the investees and investors together and establishes trust among them. Out of all the 67 million people affected by recession only a few came back solid with time because of the investments they made. The application intends to solve this problem to some extent by creating a platform where the investors, investees, stock holder, investment bankers and many field expertise can come together to help each other out. Privacy and information security is maintained among them. Perceiving experience in the market but fail to make most of it due to lack of capability to invest. There is very limited possible ways where these two people will meet and achieve each other's trust. Imagine the situation when, in recession, people relied solely on their job and when the market went down they had absolutely nothing, and then realized that there is problem. Financial magazines stated that it would happen often because of fluctuating market policies. Banking websites on the other side can reduce the chance of a dead investment but client computer. This platform enables to register with us and enjoy all the investment facilities. All the information is communicated using webservices. These web services form basic building block of our system. Data mining concept is applied in context for two purposes,the goal of data mining is to extract knowledge from a data set in a human-understandable structure.[2][3]

1. Stock/Property rates prediction: This is done using the data for 5 yrs and mining on the same to avail predictions.
2. Providing a rating (amount of trust one bestows with the system) to the investors/investees.

Proposed work would be limited to creating a platform for investors and investees to meet and have a broad discussion on their basic ideas. This would help them to get an overview of all the advancements and scope of the investments if necessary. Basic idea here is to identify the type of person (Investor, Investee, and Advertiser) and lead them accordingly. This will reduce stress from user and provide him application comfort. The application would be designed in such a way that it will be suitable to all kinds of end users; they need not be tech savvy so as to handle the application. Every module is designed in such a way that there would be scope for future advancements in all prospects so that it can also be used as a commercial application

II. RELATED WORK

A. *Stock market prediction*

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price could yield significant profit. Some believe that stock price movements are governed by the random walk hypothesis and thus are unpredictable. Others disagree and those with this

Manuscript received May 10, 2012.

D.R. Ingle*, Department of Computer Engineering Bharati Vidyapeeth College of Engineering, (e-mail: dringleus@yahoo.com). Navi Mumbai, India Mobile No 09702777927*

Dr. B.B. Meshram *Department of Computer Technology, Veer Jijamata Technical Institute, Mumbai, India.(e-mail: bbmeshram@vjti.co.in).*

viewpoint possess a myriad of methods and technologies which purportedly allow them to gain future price information.

B. SETAR and STAR

Data can be thought to fall into one of four categories as follows. Five time series: index value at open, index value at close, highest index value, lowest index value and trading volume. Fundamental factors: e.g., the price of gold, retail sales index, industrial production indices, foreign currency exchange rates. Lagged returns from the time series of interest. Technical factors: variables that are functions of one or more time series, e.g., moving averages. The standard approach to modeling stock market returns or exchange rates is to model the univariate time-series with autoregressive (AR) and moving average (MA) models. A trader can determine an appropriate number of lags for AR and ARMA models based on experience and by analyzing the time series data. Similarly, an appropriate number of regimes for SETAR (self-exciting transition AR) and STAR (smooth transition AR) models can be determined. These models are deterministic in the sense that they attempt to use mathematical equations to describe the process that generates the time-series. The advantage of these models lies in their interpretability.[4]

C. Stock Crash Prediction

Stock In recent months and weeks, we have heard many news reports about the terrible problems in the US with the failure of Enron Corporation, Global Crossing, Pacific Gas & Electric, Finova Group and Kmart. The total wipe out of assets associated with these companies is over $140 billion dollars. Comparing with the previous five major bankruptcies that took place over a 5-year period between 1987 and 1991, the total was about $20 billion dollars less. Clearly the incredible pace of globalization is leading to an ever-accelerating trend in failed companies of enormous size. It is interesting to note that Fortune Magazine estimates the search for Osama bin Laden and the Sept. 11 attacks will cost the U.S. economy $151-billion from insurance claims and increased security to protect against a repeat attack. Clearly failed companies have equal importance to a secure economy these days. The pall of the Bre-X scandal still hangs over our industry as a heavy blanket and although there have been numerous attempts to adjust the rules and make such scams a thing of the past, still there are scams continuing to take place – many based on exactly the same premises of those of the past. So this workshop is aimed at tackling the problem of mining scams head-on – to provide attendees and post-workshop readers with an insight into how these scams are promoted, how they perpetuate themselves, how we are trying to police them and prosecute those who seek to cheat others, and how we can all act to stop them dead in their tracks if we work together with a common purpose in mind.[5][6]

D. Applying Data Mining Techniques in Property Casualty Insurance

There is significant capital gain potential on mining stocks as the market begins to anticipate and discount the production of metal and the consequent earnings .The maximum appreciation is recorded if the mine is being readied for production during a period of metal prices weakness and begins pouring metal and generating earnings as prices are trending up again. In 30 years of identifying 10 companies that have qualified, all but one have equaled or exceeded their discovery highs by more than 100%! The price rises from their confirmation/disinterest lows to their production/cash flow highs have produced 300-1000% gains. Hence we can say that, One of the best times to invest in mining stocks is during the construction/pre-production period when a qualified ‘single’ ore body mining management team is going about the work of 'making mines'.

E. JD Edward’s enterprise one advanced real estate forecasting

Advanced Real Estate Forecasting gives you accurate data whenever you need it. It enables you to easily project future cash flows and pro-vide up-to-date property valuation. Real estate leasing information is used to forecast CAM, sales, percentage rent, expenses, and occupancy. Based on the most accurate, up-to-date information, you are better able to evaluate the timing of a potential sale to maximize your return on investment. And you can enhance the visibility of your future operations. By combining your existing lease data with industry-standard market assumptions about speculative leases, including leasing costs, vacancy loss, renewal probabilities, and valuation assumptions, you can quickly generate up-to-date forecasts at the property and portfolio levels.

F. Cross-Selling and Customer Loyalty in the Banking Industry

Most major financial institutions have statistics and data-mining groups. In fact, banks like Wells Fargo, Bank of America, Fleet Bank, and others have been the subject of many articles about their sophisticated data mining, and modeling of their customer’s behavior. The next question to ask is: how well do financial institutions know their customers? A study published in DM News and conducted by Deluxe Corporation found that43% of consumers surveyed said their financial service provider does not know their specific needs well at all; 60% said the offers they received were not relevant to their needs; and 39% said they did not receive offers at all. The study by Deluxe Corporation demonstrates a significant problem with data mining: the inability to leverage data-mining studies into actionable results. For example, while a bank may know that customers meeting certain criteria are likely to close their accounts, it is another matter to figure out a strategy to do something about it. One vendor that has developed a suite of products designed at integrating predictive technologies with customer interaction points is Right Point software. Other vendors are working on the same problem, particularly on the web, where predicting what a customer will best respond to is critical. Web banking companies like Security First and Broad Vision, among others, are also trying to incorporate one-to-one marketing, using predictive technologies, to their banking sites.

G. Applying Data Mining Techniques in Property Casualty Insurance

There is significant capital gain potential on mining stocks as the market begins to anticipate and discount the production of metal and the consequent earnings .The maximum appreciation is recorded if the mine is being readied for production during a period of metal prices weakness and begins pouring metal and generating earnings as prices are trending up again. In 30 years of identifying 10 companies that have qualified, all but one have equaled or exceeded their discovery highs by more than 100%! The price rises from their confirmation/disinterest lows to their production/cash flow highs have produced 300-1000% gains. Hence we can say that, One of the best times to invest in mining stocks is during the construction/pre-production period when a qualified 'single' ore body mining management team is going about the work of 'making mines'.

III. PROPOSED WORK

The main motive of the proposed system is to enhance financial communication and awareness among people. The application would be available as a webclient. Webclient is a type of a platform which is loaded onto the client computer. This platform enables to register with us and enjoy all the investment facilities. All the information is communicated using webservices. These webservices form basic building block of our system. Data mining concept is applied in context for two purposes; The goal of data mining is to extract knowledge from a data set in a human-understandable structure.

1. Stock/Property rates prediction: This is done using the data for 5 yrs and mining on the same to avail predictions.
2. Providing a rating (amount of trust one bestows with the system) to the investors/investees. [9][11][13]

Design the web based application software that can help people to get information and manage regarding their investment ideas. The system should work on databases of different data modeling schemas such as relational, object oriented etc. Read the query from end user .Create a database related to type of investor/advertiser/investee Accept further information regarding their ideas/investments. Update the same on regular basis. Maintain the integrity of the database Send e-mails and SMS for their actions on the record. Display notifications and updates related to hot news. Maintain information security by securing the personal information such as email address and mobile number. Find the most eligible investor for an investee. ETL operations like Data cleansing, Data integration, Data transformation, Data loading. .Establishing a conversation between two people who are field related via SMS and chat. Keeping a track of all the updates being done on the records. Develop the web platform to access the data from distributed system. Automated generation of investee rating.(TRUST) Generation of list of actions on a given search query.

There would be a centralized server which will provide all the clients.Investment banking using web client/web services

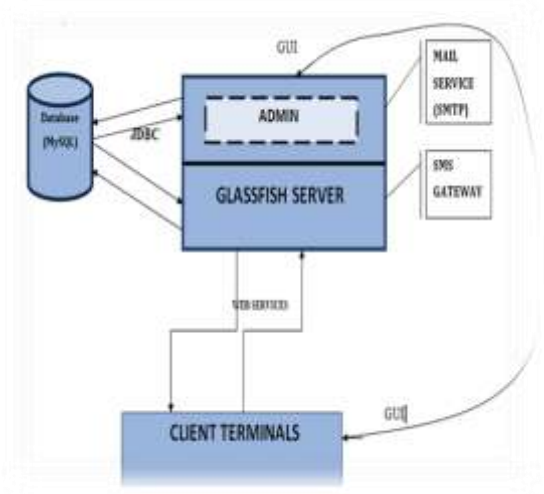


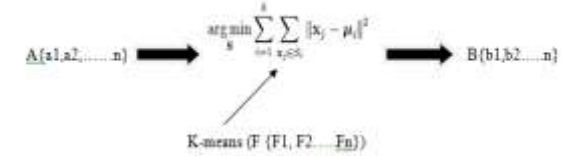
Fig. 1 Proposed Architecture

Let S be the complete system consisting of,

1. set of investor(A),
2. set of investees(B),
3. combined set consisting both of them(U)
4. a decision function(F1)
5. a set of messages(M).

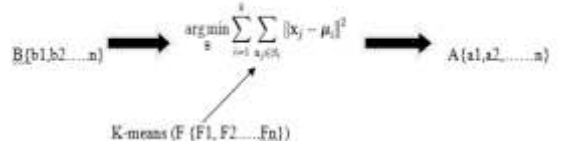
Whereas the decision function is applied in two ways

1.



In this instance, set of investors are the input to the decision function, and the set of investees are the output.

2.



It will also help the user to manage its data on one end; on the other hand it will manage data and operations on the centralized database. This database is provided with JDBC (Java Database connectivity). This will ensure the operations would be ACID. Web services would serve as means of communication; every client request would in terms generate a webservice from the platform i.e. webclient. This request in terms would be serviced by the glassfish server and again a webservice would be generated in order to reply to the client system.[12]

Set Theory :

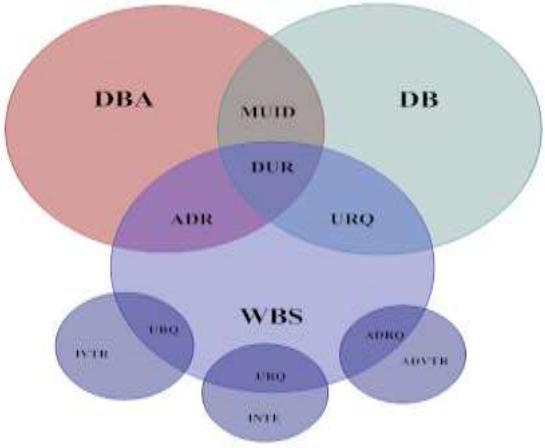


Fig.2 Set theory

IVTR : Investors ,IVTE : Investees,

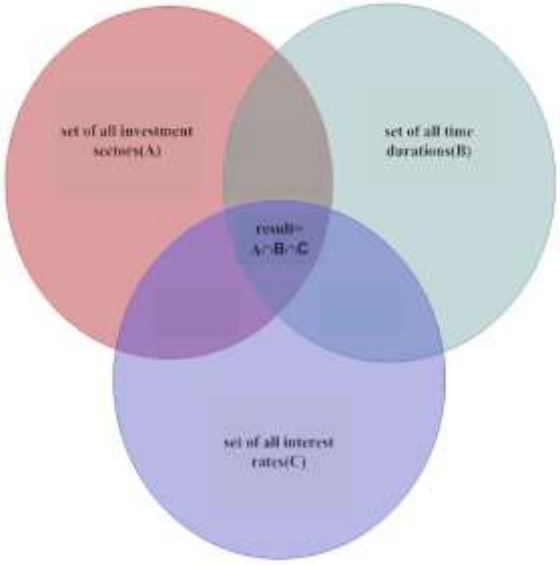
ADVTR: Advertiser

DBA: Database Administrator,

DB: Actual Database

WBS: Web services,URQ: User request

ADRQ: Advertisement Request,MUID: Manage User accounts, DUR: Database related user requests



Set theory describes a terminology to describe a given system into sets, subsets and supersets.
Our application would be mainly used where users would want to list the number of investors or investees according to a set of attributes. For example if the user wants to generate a query that will display the information of certain investors that are investing in stocks with a budget of Rs.1,00,000 for a tenure of 6 months. At this time a query containing the following information would be sent to the application server. Input set for Investor :{ inv_sector, amount, and duration} Application server would check the correctness of the arguments and generate the following query to the database. Output Set : { investees[ ], uid, description[ ]}

The first class of groups to undergo a systematic study was permutation groups. Given any set X and a collection G of bisections of X into itself (known as permutations) that is closed under compositions and inverses, G is a group acting on X. If X consists of n elements and G consists of all permutations, G is the symmetric group So, in general, any permutation group G is a subgroup of the symmetric group of X. An early construction due to Cayley exhibited any group as a permutation group, acting on itself (X = G) by means of the left regular representation. In many cases, the structure of a permutation group can be studied using the properties of its action on the corresponding set. For example, in this way one proves that for $n \geq 5$,the alternating group An is simple, i.e. does not admit any proper normal subgroups. This fact plays a key role in the impossibility of solving a general algebraic equation of degree $n \geq 5$ in radicals.

Reason

Mining problem can be thought of as a larger problem been broken down into smaller sub problems.

Example: We can apply mining on initial records to get some result or we can re-apply mining on prey results to get more advanced relations.

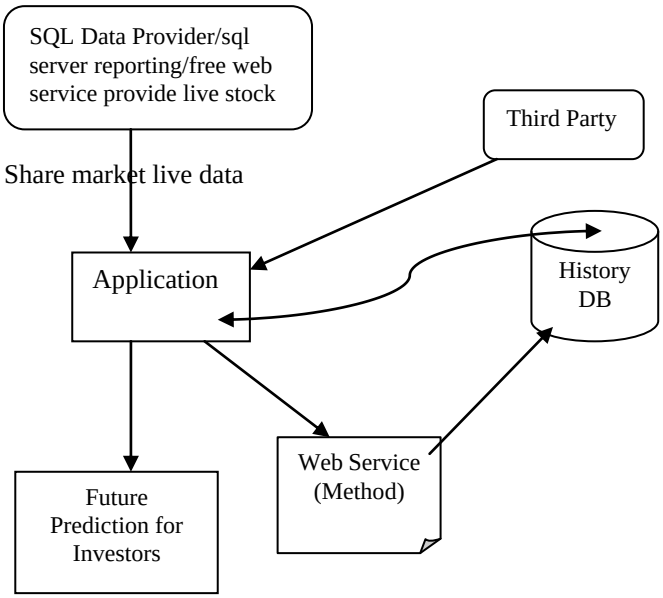
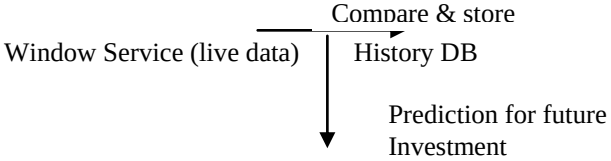


Fig. 3 Proposed Solutions



We can get live stock data by using SQL Data Provider/sql server reporting/free web service provide live stock and simultaneously it is stored in History DB. By using data mining algorithm mining like K – means algorithm we can cluster the data in the groups showing drop & rise conditions of share markets in the particular time period by using decision tree algorithm we can predict the Investee's future whether it is fall or rise . By using this proposed solution we can predict that where the investor is in benefit or loss in the share market.

A. K- Means Algorithm

The k-means algorithm is a simple iterative method to partition a given dataset into a user specified number of clusters, k. This algorithm has been discovered by several researchers across different disciplines.

This algorithm would be used in the project for choosing the best investor for a queried investor and vice-versa. In statistics and data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This results into a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard), however there are efficient heuristic algorithms that are commonly employed that converge fast to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data, however k-means clustering tends to find clusters of comparable spatial extend, while the expectation-maximization mechanism allows clusters to have different shapes.

Standard algorithm:The most common algorithm uses an iterative refinement technique. Due to its ubiquity it is often called the k-means algorithm; it is also referred to as Lloyd's algorithm, particularly in the computer science community.

Given an initial set of k means m1(1),…,mk(1) (see below), the algorithm proceeds by alternating between two steps.

Assignment step: Assign each observation to the cluster with the closest mean (i.e. partition the observations according to the Voronoi diagram generated by the means).

Update step: Calculate the new means to be

$$S_i^{(t)} = \left\{ \mathbf{x}_j : \left\| \mathbf{x}_j - \mathbf{m}_i^{(t)} \right\| \le \left\| \mathbf{x}_j - \mathbf{m}_{i^*}^{(t)} \right\| \text{ for all } i = 1, \ldots, k \right\}$$

Update step: Calculate the new means to be the centroid of the observations in the cluster.

$$\mathbf{m}_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{\mathbf{x}_j \in S_i^{(t)}} \mathbf{x}_j$$

The algorithm is deemed to have converged when the assignments no longer change.

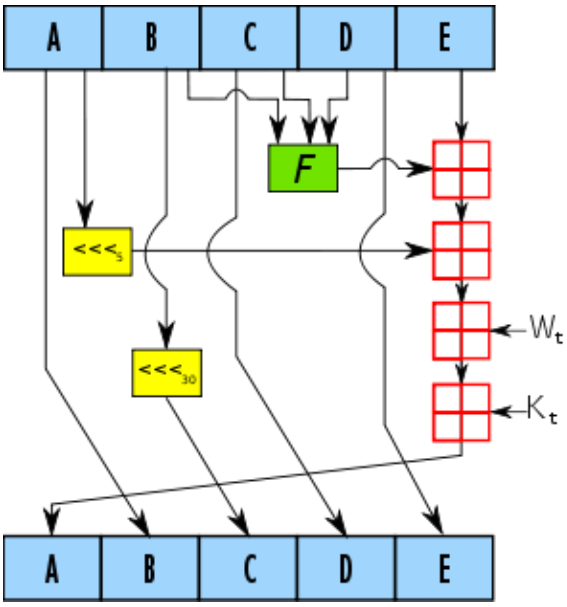
Commonly used initialization methods are Forgy and Random Partition. The Forgy method randomly chooses k observations from the data set and uses these as the initial means. The Random Partition method first randomly assigns a cluster to each observation and then proceeds to the Update step, thus computing the initial means to be the centroid of the cluster's randomly assigned points. The Forgy method tends to spread the initial means out, while Random Partition places all of them close to the center of the data set. According to Hamerly et al., the Random Partition method is generally preferable.Security Implementation: **SHA-1**

This algorithm or rather function would be helpful in our project to maintain the security of information (contact details and passwords), so that even our integral parts such as administrator wouldn't be able to view such information.

Description:

SHA-1 produces a 160-bit message digest based on principles similar to those used by Ronald L. Rivest of MIT in the design of the MD4 and MD5 message digest algorithms, but has a more conservative design.

The original specification of the algorithm was published in 1993 as the Secure Hash Standard, FIPS PUB 180, by US government standards agency NIST (National Institute of Standards and Technology). This version is now often referred to as SHA-0.



Kt is the round constant of round t;

SHA-1 differs from SHA-0 only by a single bitwise rotation in the message schedule of its compression function; this was done, according to NSA, to correct a flaw in the original algorithm which reduced its cryptographic security. However, NSA did not provide any further explanation or identify the flaw that was corrected. Weaknesses have subsequently been reported in both SHA-0 and SHA-1. SHA-1 appears to provide greater resistance to attacks, supporting the NSA's assertion that the change increased the security.

B. Decision Tree Algorithm

Type of tree-diagram used in determining the optimum course of action, in situations having several possible alternatives with uncertain outcomes. The resulting chart or diagram (which looks like a cluster of tree branches) displays the structure of a particular decision, and the

interrelationships and interplay between different alternatives, decisions, and possible outcomes.

A decision tree presents a decision procedure to determine the class of a given instance. Each node of the tree specifies either a class name or a specific test that partitions the space of instances at the node according to the possible outcomes of the test. Each subset of the partition corresponds to a classification sub problem for that subspace of the instances, which is solved by a sub tree. A decision tree can be seen as a divide-and-conquer strategy for object classification. Formally, a decision tree can be defined to be either:

a). a leaf node (or answer node) that contains a class name, or b). a non-leaf node (or decision node) that contains an attribute test with a branch to another decision tree for each possible value of the attribute [4]. Nowadays the most popular decision tree algorithms include: CHAID (Chi-squared automatic interaction detection), QUEST (quick, unbiased, and efficient statistical tree) and CART (Classification and regression tree).

C. Interaction with system

Investor, Investee and Third party like advertiser can interact with the proposed system. They have has a relationship

Investor can invest their capitals with the help of our proposed system by creating their account. System is able to provide different options for secure investment.

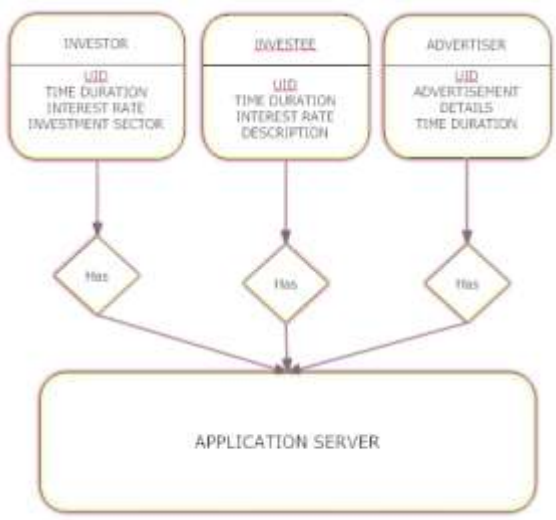


Fig.4 Interaction with System

Investors can invest by considering the term like time duration, interest rate, investment sectors ect. Investee will provide all details of their service.

D. Web Services

We have made use of web service oriented architecture because of which our design had been simplified. This has given us a broad perspective of the client server communication. End user could be any wired, wireless workstation (laptop, mobile, computer, PDA). A Hyper Text transfer Protocol is used as a means of communication to the application server, thus providing scope for incoming requests Business Partner or other systems are nothing but the data administrator usage in the application. We have made use of the EJB (Enterprise JavaBeans) in order to create modular structure on the server side. Inter-block communication would be provided using web services as those are the fastest available in the market. Database serves as a bottleneck for providing query responses.

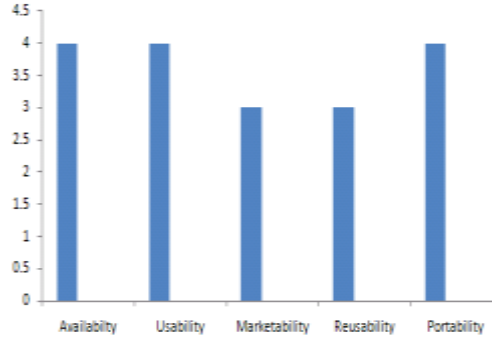
E. Quality Attributes

Proposed system would be available to the end user at every moment. End user need not be aware of complex market terms.

User has a control over any type of postings, so the same user can take advantage of posting and viewing multiple queries.

This system is very likely to adapt in the market due to high need of analytic systems. every component can be reused in a high level stock, property and business analysis.

It can run effectively on every environment with JRE.



Future:

- Stock prediction (To determine the stock values in future).
- Stock Broker profiling (To determine the amount of trust a broker has achieved)
- Business Intelligence Systems (To help an organization to make good business decisions)
- REC profiling and property rates prediction (To determine the rise and fall in property rates).

IV. CONCLUSION

The main purpose of this system is to provide user with the investment facilities and detect stock and property rates; this is achieved with data mining like K- Means Algorithm and Decision Tree Algorithm. These algorithms are very efficient in clustering the history data of share market & making appropriate decision when to invest. It is best on prevision/ history data of share market this will be helpful in protecting the future of investors. System would be faster and cost effective in terms of execution and generating effective results due to use of web services. It would be

absolutely free of cost to everyone, except the advertisers. Proposed sytem doesn't require any special maintenance program as backups and updation would be automatic at server end. User privacy is a major advantage as user would be ableto get the appropriate listings keeping their privacy (mobile number and e-mail address) intact. Marketers of a software company may advertise about their new software to consumers who have a lot of software purchasing history.

REFERENCES

[1] Judith M. Myerson "New Directions in Internet Management" Enterprise Systems Integration, 2nd Edition and the Complete Book of Middleware articles

[2] Adam Kie˙zun MIT akiezun@csail.mit.edu Philip J. Guo Stanford University pg@cs.stanford.edu Karthick Jayaraman Syracuse University kjayaram@syr.edu Michael D. Ernst University of Washington mernst@cs.washington.edu

[3] Hazeline U. Asuncion and Arthur U. Asuncion"Software Traceability with Topic Modeling"

[4] Hazeline U. Asuncion. "An End-To-End Industrial Software Traceability Tool".

[5] By Huzefa Kagdi, Jonathan I. Maletic and Bonita Sharif "Mining Software Repositories for Traceability"..

[6] Jay Tomlin "Secure Gateway 3.0 for Presentation Server".

[7] Edieal J. Pinker, Abraham Seidmann and Yaniv Vakra "Managing Online Auctions "

[8] Layne Lathram "Online auction sites and trademark infrigeability".

[9] Dunwei Wen and Jiangqiang Xiong" SMS Based Querying ".

[10] Hazeline U. AsuncionInstitute"An End-To-End Industrial Software Traceability Tool" for Software Research University of California, Irvine Irvine, California 92697+1-949-824-2703 hasuncion@ics.uci.edu, Frédéric François Wonderware Corporation26561 Rancho Parkway SouthLake Forest, California 92630+1-949-303-8388taylor@ics.uci.edu

[11] Huzefa Kagdi, Jonathan I. Maletic, "Mining Software Repositories for Traceability", Links Bonita Sharif Department of Computer Science Kent State UniversityKent Ohio 44242 {hkagdi, jmaletic, bsimoes}@cs.kent.edu

[12] Andrew W. Moore "K-means and Hierarchical Clustering"Associate Professor School of Computer Science Carnegie Mellon University www.cs.cmu.edu/~awm awm@cs.cmu.edu 412-268-7599

[13] Edieal J. Pinker • Abraham Seidmann • Yaniv Vakrat W. "Managing Online Auctions: Current Business and Research Issues" E. Simon Graduate School of Business Administration, University of Rochester,Rochester, New York 14627

[14] Judith M. Myerson,"Web Service Architecture" Published by Tect, 29 South LaSalle St. Suite 520 Chicago Illinois 60603 USA

[15] Edieal J. Pinker • Abraham Seidmann • Yaniv Vakrat W. E. Simon "Managing Online Auctions: Current Business and Research Issues" Graduate School of Business Administration, University of Rochester, Rochester, New York 14627

[16] Andrew W. Moore , "K-means and Hierarchical Clustering" ,Associate Professor School of Computer Science Carnegie Mellon University www.cs.cmu.edu/~awm awm@cs.cmu.edu 412-268-7599

[17] School Street N. Chelmsford," Web Services . Ten Examples of Real World ROI"MA 01863-2109

[18] XindongWu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh ,"Top 10 algorithms in data mining"
.

First Author:

Mr. D.R. Ingle (ISTE LM'2004) is Associate Professor of Computer Engineerig Department **at** Bharati Vidyapeeth College of Engineerig, NaviMumbai, Maharastra state, India. He received bachelor degree, and Master degree in computer engineering. He has participated in more than 10 refresher courses to meet the needs of current technology. He has contributed more than 20 research papers at national, International Journals. He is life member of Indian Society of Technical Education.His area of interest are in Databases, intelligent Systems, and Web Engineering.

Second Author:

Dr. B.B.Meshram (CSI LM'95, IE '95) is Professor and head of Computer Technology Department at VJTI, Matunga, Mumbai, Maharastra state, India. He received bachelor degree, Master degree and doctoral degree in computer engineering. He has participated in more than 30 refresher courses to meet the needs of current technology. He has chair more than 15 AICTE STTP Programs and conferences. He has received the appreciation for lecture at Manchester and Cardip University, UK. He has contributed more than 200 research papers at national, International Journals. He is life member of computer society of India and Institute of Engineers. His current research interests are in Databases, data warehousing, data mining, intelligent Systems, Web Engineering and network security.