Grade Prediction & Analysis Methods for University System

A. Yasara Dissanayake

Abstract— **Here I describe system of analyzing the students marks in several weeks in semester and method of comparing them with other students and system of predicting the future result based on previous years. The aim of this system is to the lecturer to improve the grades of the students at the final exam using Hidden Markov Model and 3 regression methods which are Polynomial Regression, Exponential Weighted Average Regression and Exponential Moving Average Regression by combining using Weighted Regression. This article shows the final output of the system and how they are compatible with real results using unit testing.**

I. INTRODUCTION

"What do I need to make on my remaining tests and assignments in order to get an A+ in final exam, given the grades that I've already made so far?" My University grade predictor was created to answer that question! I created this application because every semester, students crank up their calculators in pursuit of the answer to this elusive question on their final grade. It's easy to make mistakes, though, especially if you have tests and assignments that are worth different percentages of the total grade.

At the moment even lectures have their students result but they cannot compare it with others and also they cannot analyze them and cannot get idea about successiveness or fail of their studies before the examinations. This situation affects the achieving lower grade of some student. And also when lecturer have to give feedback on student when the situation like company interviews its very difficult to get overview of the student in current system. By this application since he can get early warning before closer to examination, he can upgrade the result of batch by giving early warnings to the academically weak student based on future predicted result. And also he can uplift the standard of department to the industry by giving correct feedback on students

For the prediction of grades this system analyzes the student studying pattern as how the auto scalier predict the usage of the system using different regression models.Regression models are often constructed based on certain conditions that must be verified for the model to fit the data well, and to be able to predict accurately.Here in the system mainly used polynomial regression and it is a form of linear regression in which the relationship between the independent variable x and the dependent variable y is modeled as an nth degree polynomial. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y — x), and has been used to describe nonlinear phenomena is the the main reason for selecting that model. And also this system combines the typically related subjects predict ones result from other related subject using Hidden Markov mode.This model useful to imagine an generating a sequences of results and get the mean value of that result and predict final result. When we visit a state, we emit a residue from the state's emission probability distribution. Then, we choose which state to visit next according to the state's transition probability distribution. The model thus generates two strings of information. One is the underlying state path (the labels), as we transition from state to state. The other is the observed sequence (the related subject results), each residue being emitted from one state in the state path.

System is implemented as a standalone system as since because of the legal condition at my targeted university system; which is University of Moratuwa, it is not allowed to publish marks of other students in public. And the main reason for using python for this system is it is best language for scientific system like this to improve efficiency and accuracy of the mathematical models which are use for prediction.

The basic purpose of this application is improving the grades of students at the examination. According to current situation student cant have any idea about his grades until he write the exam. By this application student can have early warning by his lecturers about their low performance and students has time to improve it before he attempt to any examination. When at the interview since student can have more correct feedback regarding their performance the standard of department may increase. Work load of a lecturers also reduce since number of files which they have to maintain are decrease

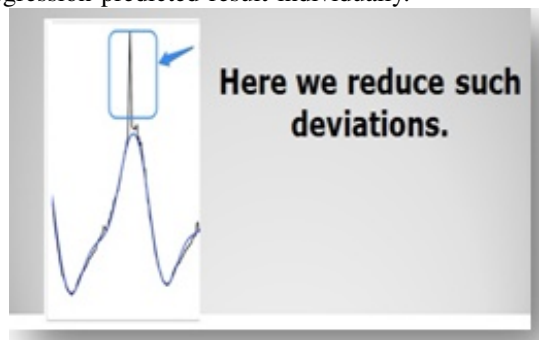
This application contain 4 algorithms which is mentioned in abstract including Markov Model which gives the comparison based on other subject performance and considering their accuracy again assign weights for them and do linear regressions again the system consists of graphical method to show outputs such as comparison of result to make easy of understanding. Used approach of the system is lecture can use text files to configuration and data insertion functions. When at the configuration level password and user-names keeps backups in database too and they are in encrypted format when writing to the file. System use memento design pattern to restore the password when user forgot the password or when system has hacked. To improve the performance configuration of Markov Model can be done at the beginning by entering the grades to the text files and then the transmission and emission probabilities are calculated and keep in saved files to use later(If it calculate each time of running CPU power is getting increase). Warnings are

automatically handles by triggers when the current predicted marks become below 35 and displayed it in the separate window at each time including marks and subject.

Apart form the main functionality this system contains graphical representation which make user friendly to analyze and comparing marks of the students each other, encryption(64BaseCode system) to increase the security and privacy of data, design patterns(Singleton and Memento), database system including triggers and event handling improve the quality of system.

II. LITERATURE REVIEW

The basic idea for this system is get from the existing auto scalar called Scryer Netflixs Predictive Auto Scalar [1] which is introduce for prediction purpose of cluster of data points. It has given that every data cluster has particular periodic pattern. This system contain concept for student learning pattern. That means every student has pattern of studying like some student work on daily and some of them work only at near the exam. Thats why this system contains used regression for one of the prediction part. But since student dont have linear pattern of studying this system contain used polynomial regression instead of linear regression which is originally suggested. For that system contain method which is currently on Apache Stratos auto scalar [2]. And also for make more accurate prediction by removing extraordinary variation (Figure 2.1) this system contain Exponentially Weighted Moving Average [4] since for one prediction method I use assignment there can be more variations than student's expected value. Apart from that method this system use 2 other regression methods because to to predict more accurate patter and use weighted regression by testing results with actual result by using regression predicted result individually.



(Fig-2.1)

As a second type of prediction this system contains Hidden Markov Model [3] which gives relational prediction on the series of observation. A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. A HMM can be presented as the simplest dynamic Bayesian network. In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states. Note that the adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a 'hidden' Markov model even if these parameters are known exactly.This system assigned those observations as previous examination result and output as a predicted result. Since that have 2 probabilities which call emission and transition It gives more relation output result.

Advantage of this system apart from the main system functionalities are since it has UI which are more closer to internationally accepted standard system can use even a person who has not knowledge in IT. And also this system displays the outputs in graphical format it is easy to analyze and compare the results each other easily. Security and the privacy of the data are well established using 64basecode encoding system and using Singleton design pattern which helps to protect system from multi threaded attacks. For make the more user friendliness in the system this system contain Memento design pattern which help to restore to the to the previous state when user forget his password and user name. All the files are protected in the data base except which are using as text files for more security.

In currently there is some grade prediction software like Learningplusuk and is provide Allows the user to determine subject/course level predictions at Level 3,Calculates the average point scores for GCSE grades,Provides predictions at the 50th to 80th percentile based on the complete national dataset and Enables predictions for entire cohorts. But i does not have logical relationship between subjects. And also in grade predictor of University of Aliban Only courses graded A-E will factor into to that GPA and no transfer coursework will factor into your UAlbany GPA like in this system uses both CA marks and final grades.

There is another system in our batch which do the same functionality such as prediction. But is has used the method such gives the prediction by considering previous batch similar performance student mark. Since in this system student individual history as well as probability value considering whole batch when calculating prediction this system may gives more accurate prediction than that.

III. SYSTEM MODELS

A. *System Requirement*

i. Functional requirements of the system

As a personalization requirement this system Provide facility to store marks in database within that particular academic time. After time period they can delete tables or can store it for next semesters.Database is automated using triggers and event handling which is easily clear the entries after the one academic semester.

As a data reporting this system data recorded on the database can be represented using graphs showing the progress of the students over time. It will show the variation of the marks over time as well as different students.This

system has UI which are more closer to internationally accepted standard system can use even a person who has not knowledge in IT. And also this system displays the outputs in graphical format it is easy to analyze and compare the results each other easily.

As a data entry requirement this system provide facility which lectures can manually enter the data to the system if they are not already in database. lecturer can copy and past marks data to the test files and then files are scanned by system and sore that values to the data base. this make lecturers work more easy.

Security and the privacy of the data are well established using 64basecode encoding system and using Singleton design pattern which helps to protect system from multi threaded attacks. For make the more user friendliness in the system this system contain Memento design pattern which help to restore to the to the previous state when user forget his password and user name. All the files are protected in the data base except which are using as text files for more security.

ii. Non-functional requirements of the system

Functionality-System is acceptably efficient when calculating output.User should not wait for a long time to and display those outputs.System occupy low memory specially because this is for individual user, if this system takes high memory then users may refuse to use this system. Since the some of the models are created and write those parameters like Markov model transition and emission probabilities at the system configuration level, this system gives outputs taking less time.

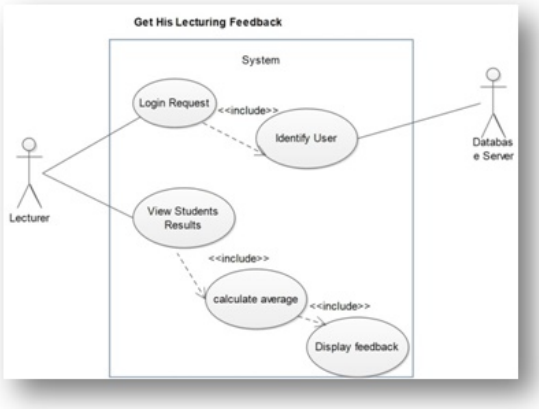
Usability-System has user friendly and very low in training time.Since this system is for both school and university lectures we could not expect to have good knowledge in software handling from them. Therefore all guidelines should be given and UI should be very clear for use. Each and every functional option contain separate help window which has guidelines for inputs and usage.

Reliability-System is well accurate When calculating the prediction the future result. And also it needed when taking the decision on course successiveness or fail in particular course.By looking at the unit test output it can be see that the system has considerable reliable level. And also this system can check weather the system is hacked and changed the password.

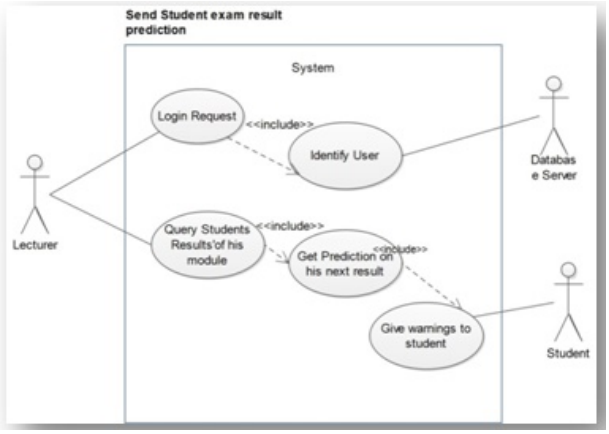
Performance-Response time is very high .Resource utilization low since this system mostly uses the text files than database tables. Only the important data are stored in data base. Resource utilization, such as memory, disk, communications etc. is very low since this system use in single computer. Since the some of the models are created and write those parameters like Markov model transition and emission probabilities at the system configuration level, this system gives outputs taking less time. And also triggers in the database help part of the calculation.

Support ability-System should provide well guideline.Since this is for student which has no good knowledge on software system should be user friendly.This system has UI which are more closer to internationally accepted standard system can use even a person who has not knowledge in IT. And also this system displays the outputs in graphical format it is easy to analyze and compare the results each other easily.

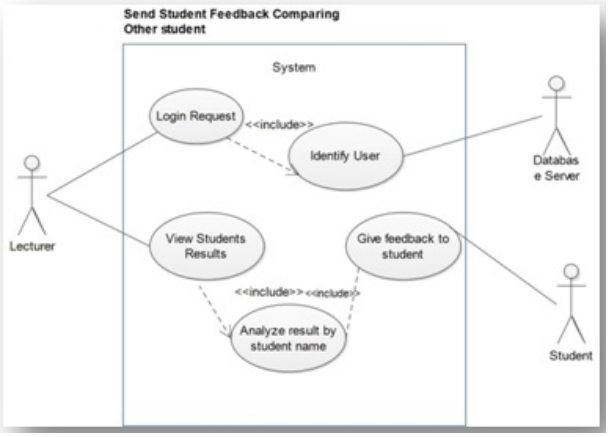
Main use-case diagrams of the system



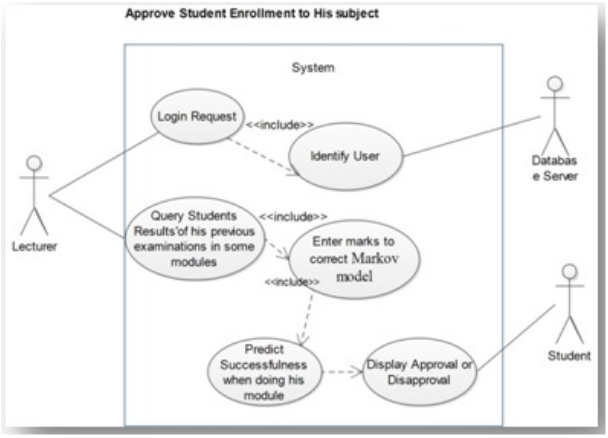
(Fig-3.1) This use case (figure 3.1)is about when student or industry people came and ask about student the lecture can view his progress and can give more accurate feedback



(Fig-3.2) This use case (figure 3.2)is about when lecture can view low grade student and warn them before exam considering prediction



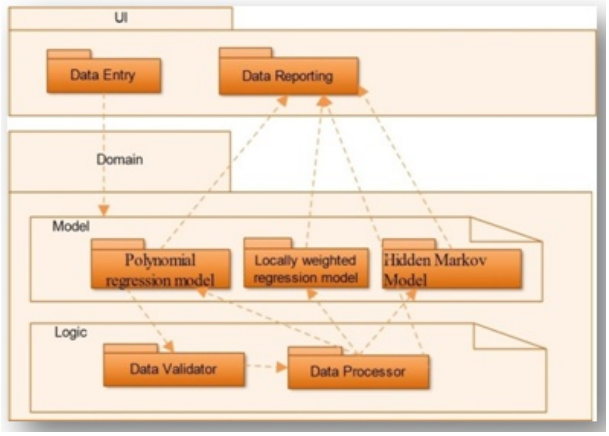
(Fig-3.3) This use case (figure 3.3)is about Lecture view the progress of marks in all batch



(Fig-3.4) This use case (figure 3.4)is about when lecture can get prediction when allow student to follow his course based on previous subject result

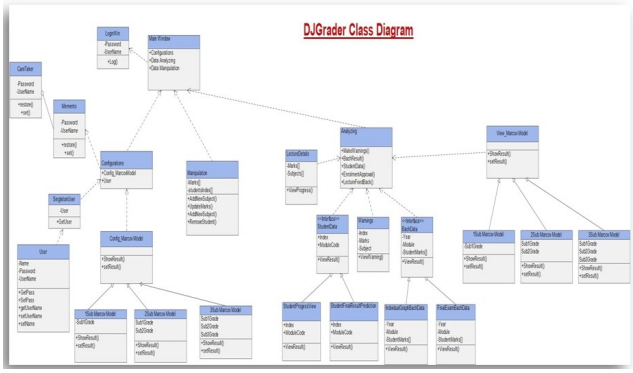
B. System Design

The following UML 3-tier architecture diagram(Figure 3.5) shows the main layers of the architecture of the application, the packages within each one of them and interactions between these packages.



(Fig-3.5)

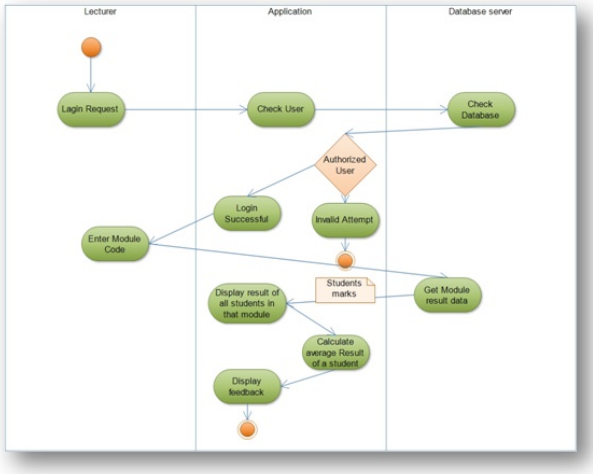
The class diagram of the system and describe the classes and associations are shown below(figure 3.6).



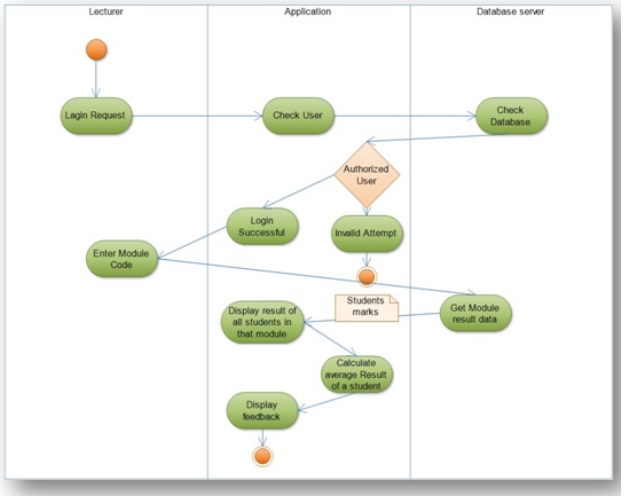
(Fig-3.6)

In this class diagram memento and care taker are the classes which is used for memento design patter. care taker class hold the previous stage of the system which can be used in when user forgot his new password. Singleton user class is the the class which use for singleton design patter which helps to protect system from multi threaded attacks and create and single user at one time. Other classes are created for each functionalities.

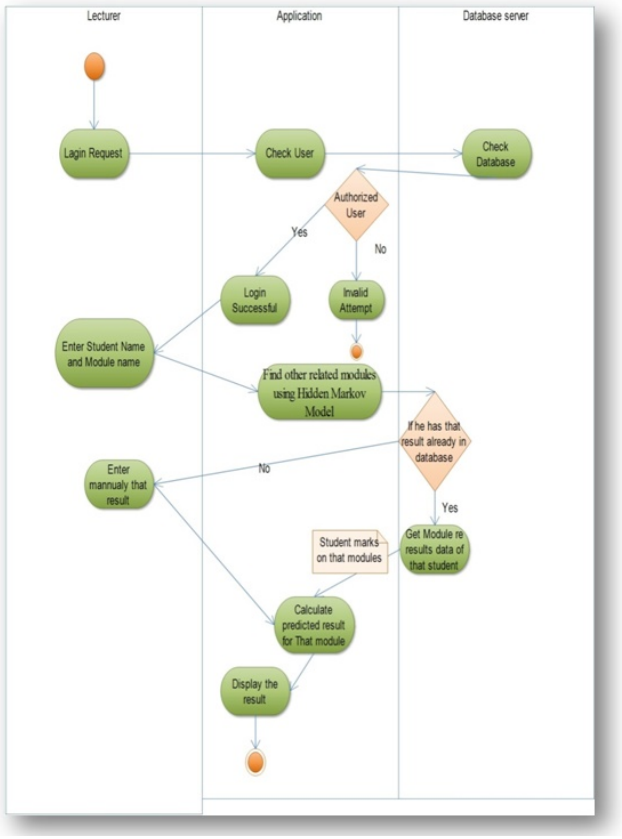
Activity diagram of the system



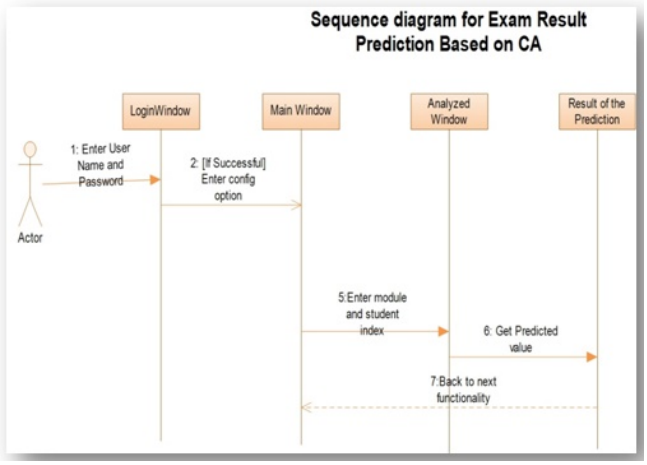
(Fig-3.7) This activity diagram (figure 3.7)is about when lecture can view low grade student and warn them before exam considering prediction



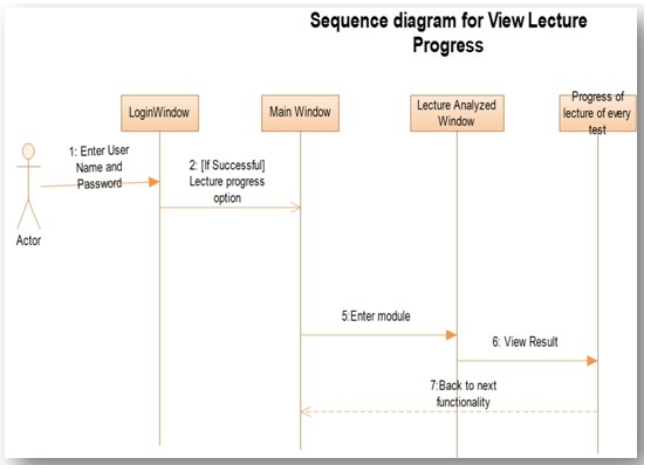
(Fig-3.8) This activity diagram (figure 3.8)is about Lecture view the progress of marks in all batch



(Fig-3.9) This activity Diagram (figure 3.9)is about when lecture can get prediction when allow student to follow his course based on previous subject result

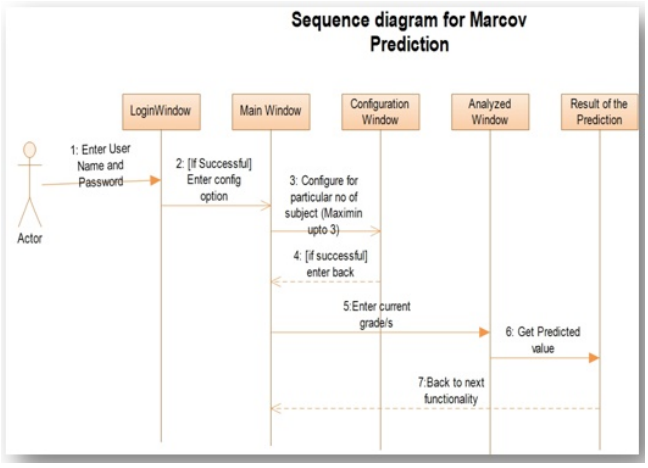


(Fig-3.11)



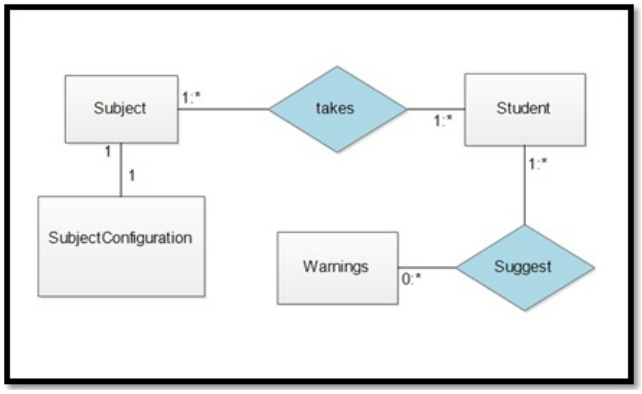
(Fig-3.12)

Main sequence of the system



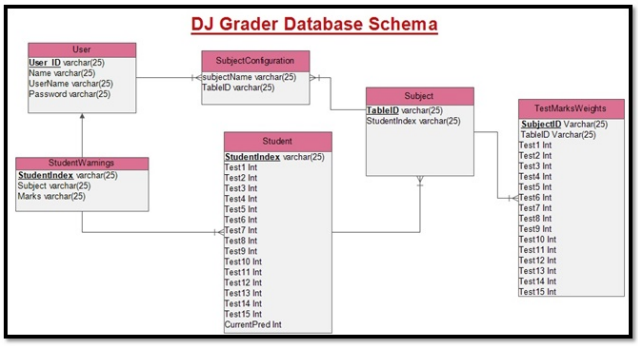
(Fig-3.10)

Database Design -ER Diagram of the system database



(Fig-3.13)

Database Design -Database schema of the system database



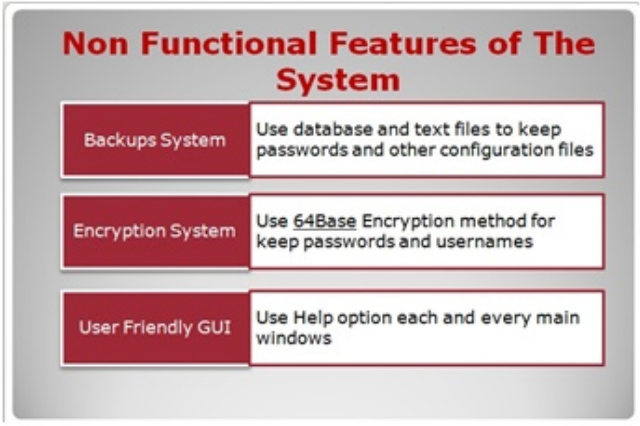
(Fig-3.14)

Here system has used triggers for automatically updated the warnings table when student achieved marks below than 35. And by using event that table entries are erased by after 1 week. Weights table contains the weights of the each test which used for predicting final based on CA value because not ever tests never carry equal marks.

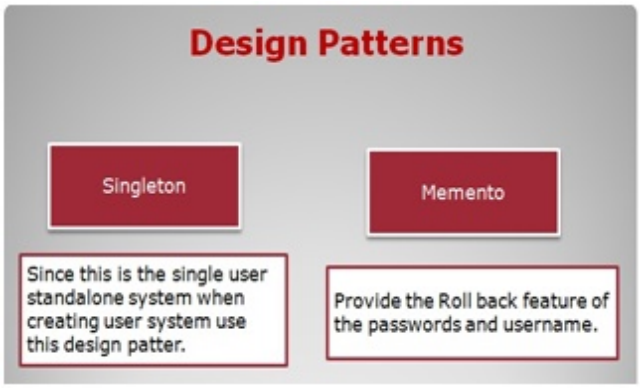
IV. SYSTEM IMPLIMENTATION

A. Implementation Procedure

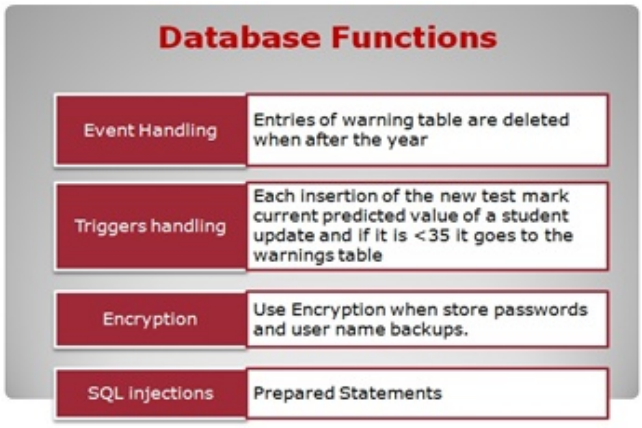
Since this is a scientific system system has used Python as a language to improve the accuracy and efficiency of the system. And also this system is implemented as standalone system since the publishing marks of the student mot legal process. As a IDE system has been used JetBran PyCharm IDE and for the data base system has been used MySql and Wamp server. In main functionality sections including calculation of probability, calculation of weights, exponential moving regression, weight regression, implementing all data manipulation functions, system configuration functions, implementing the design patterns etc, parts are done manual coding. system has been used pyplot frameworks for displaying graphs. In functional development system has been used scikit learning curve fitting library for generate polynomial function and for Hidden Makov model system has been used hmm library. Other than the main functions below (Figure 4.1 and 4.2) mention functions are implemented within the system for improving the quality of the system.



(Fig-4.1)



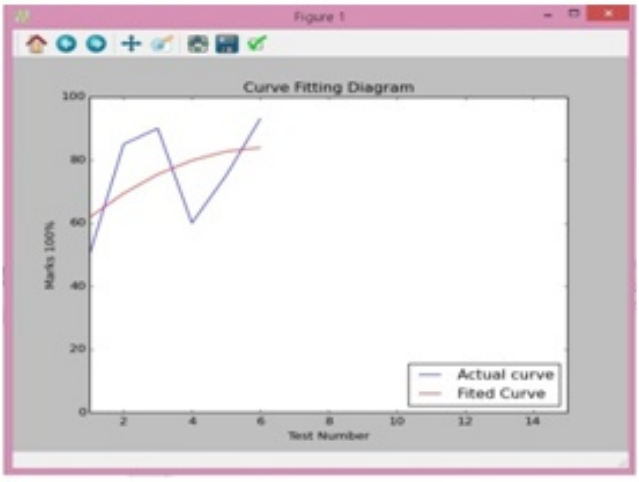
(Fig-4.2)



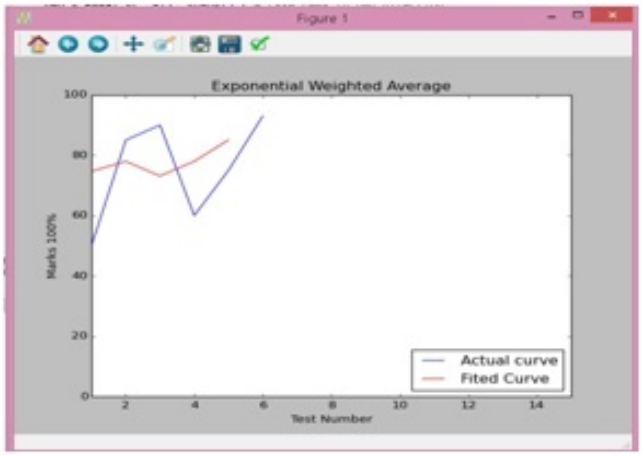
(Fig-4.3)

For each functional implementation and GUI implementation system has tested functionality by unit testing as mentioned in testing section. At the data base implementation part system has used following techniques for the system make more secure and reliable.

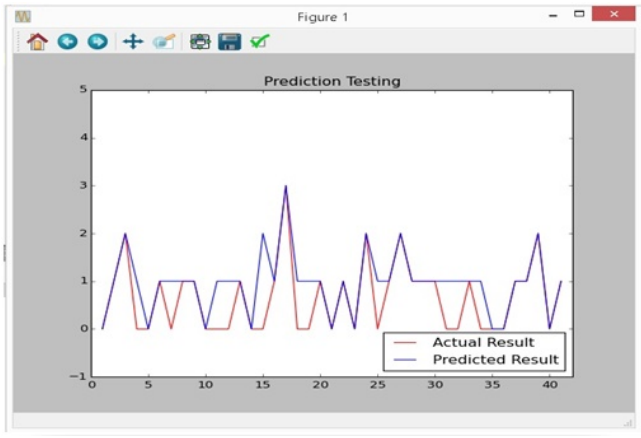
Unit testing on individual algorithms on student one current student's marks based on CA when prediction in test cases



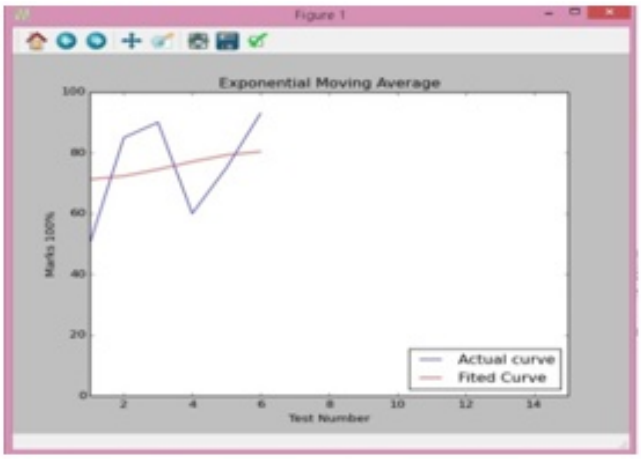
(Fig-4.4) Curve Fitting



(Fig-4.5) Exponential Weighted Average

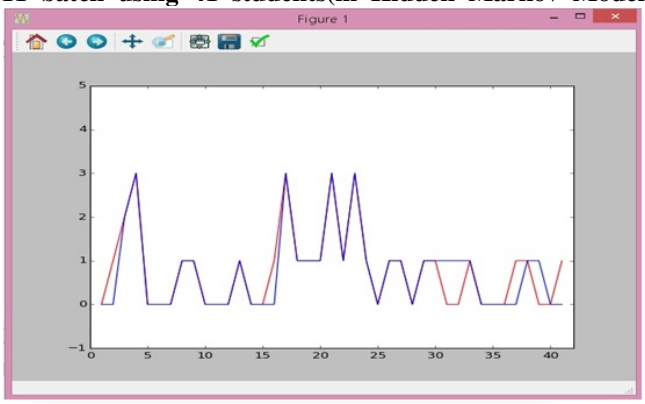


(Fig-4.8)



(Fig-4.6) Exponential Moving Average Regression

Accuracy Testing By Unit Testing in OOP and OOSD modules Related to Software Engineering module in 11 batch using 41 students(in Hidden Markov Model)



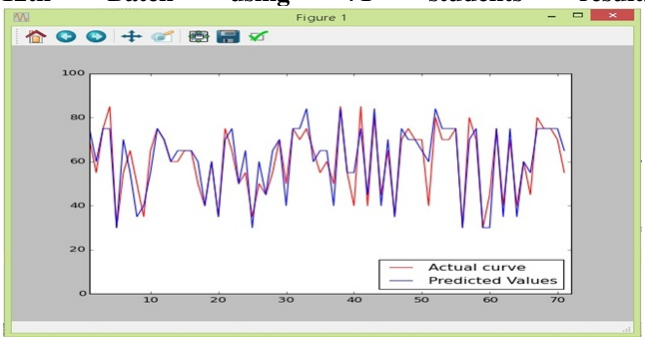
(Fig-4.9)

UI Testing on Function of button click

```
from mock import patch
import win_HomeWinl
import unittest
import Tkinter
'''
This is the test function for the button in home wiondow wjich call OS.System funcrion
'''
class HelloCallBack4TestCase_system(unittest.TestCase):
    def setUp(self):
        self.root = Tkinter.Tk()
        self.MainWindow = win_HomeWinl.MainWindow(self.root)

    def tearDown(self):
        self.root.quit()

    @patch('os.system')
    def test_helloCallBack_os_system(self, mock_tk_os_system):
        self.MainWindow.helloCallBack4()
        self.assertTrue(mock_tk_os_system.called)
        print "mock_tk_os_system_call_True"

if __name__ == '__main__':
    unittest.main()
```

(Fig-4.7)

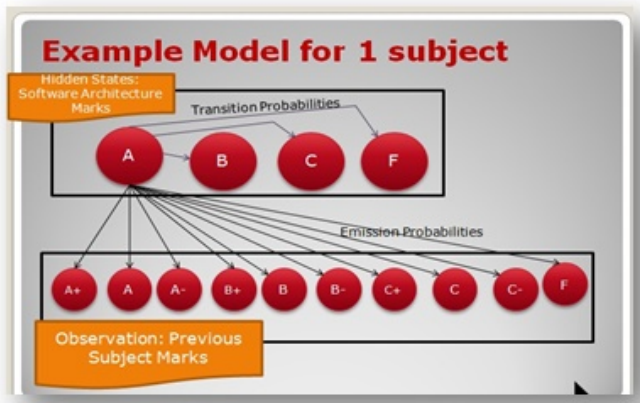
In Data Structure module of a 12th Batch using 71 students results



(Fig-4.10)

Accuracy Testing By Unit Testing in OOP module Related to Software Engineering module in 11 batch using 41 students(in Hidden Markov Model)

Other than those things, by creating different privileged users test the security of the system. When after integration all sub systems to the main system finally tested performance and security as mentioned in testing section.

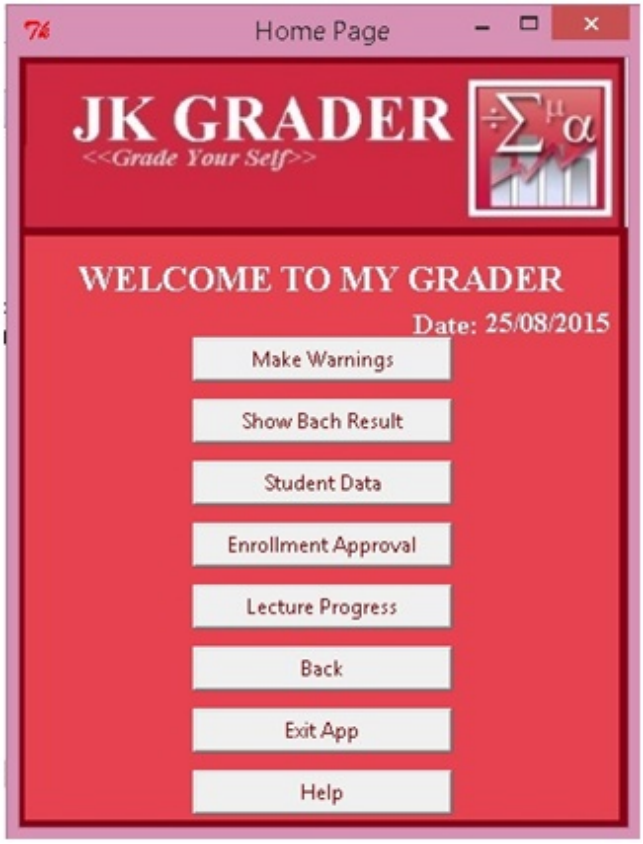


(Fig-4.11)

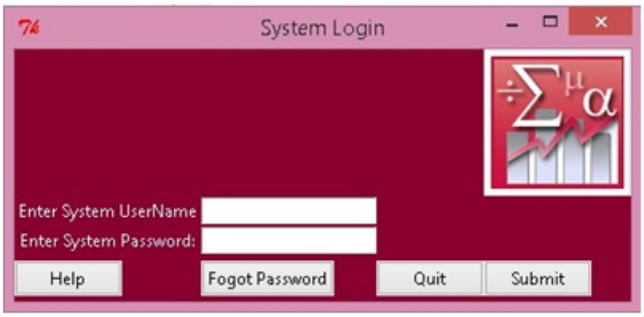
System has considered previous batch final results of subjects like OOP, OOSD and Software engineering result of 11 Batch in our university for calculating probability values. For regression. System has used CA marks of Communication Skills, Data Structures and algorithms, Computer architecture from the lecturers. system has used above data for accuracy testing purposes. system has could get the final grades (A+, A, B,) values of the final grades then system has converted them to average value of them (Ex: For A+=85+ (100-85)/2= 92.5) and use for prediction and when in the database system has used integer form by round off that values.

B. Main Interfaces

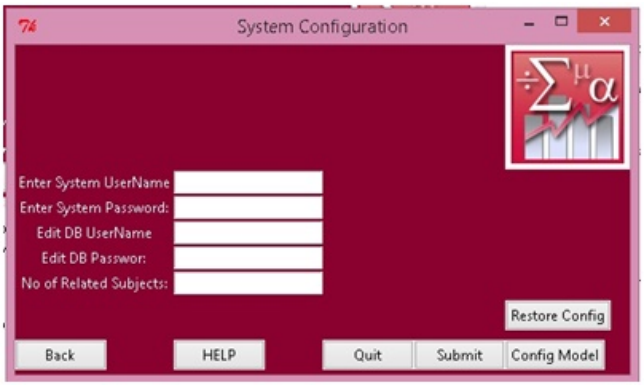
Main UI Windows



(Fig-4.12) Home Window which contain all the main functional options

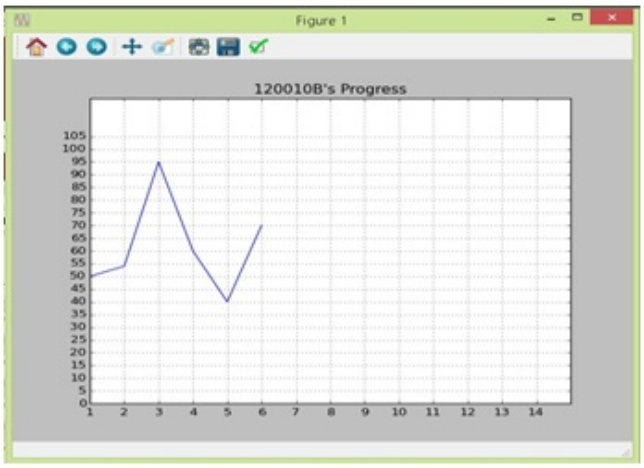


(Fig-4.13) Login Window is the very 1st window which can be login 1st time at a default user name and password

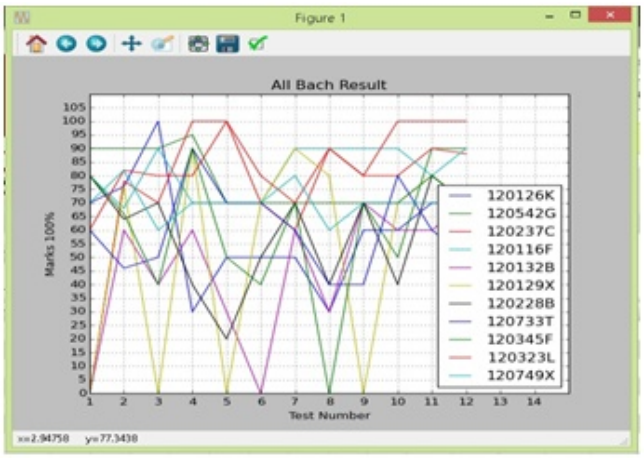


(Fig-4.14) Configuration Window. This window set the palatially calculated model parameter as well as restoring facility of the system.

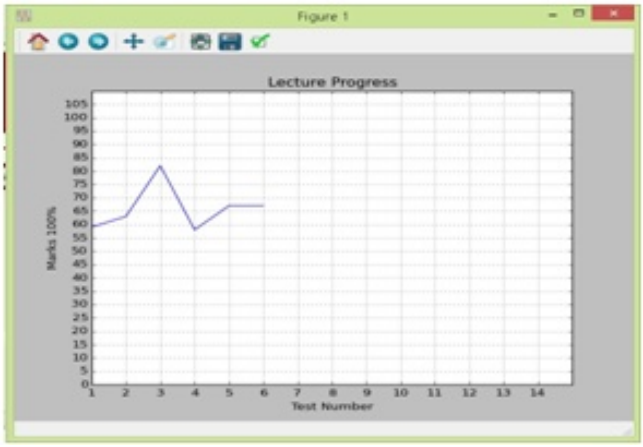
Main Output Windows



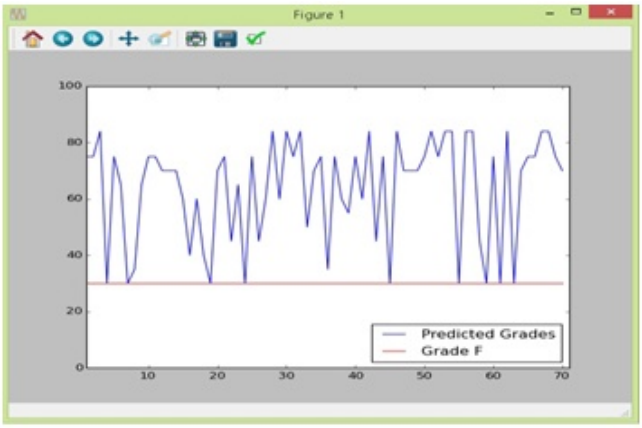
(Fig-4.15) This window View the Lecture Progress of the lecturer in several week times.



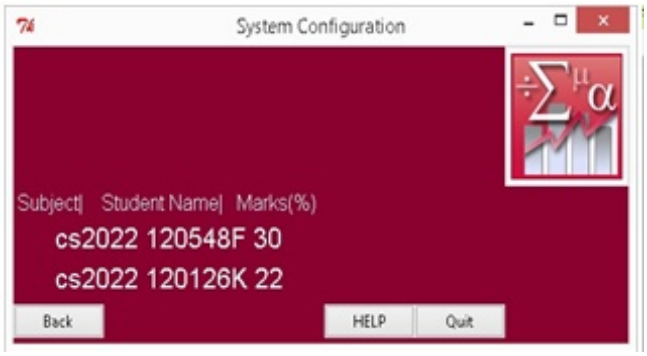
(Fig-4.16) Compared to whole batch. Using this widow lecturer can identify the weak students as it display the student name separately.



(Fig-4.17) Individual Student progress. Using this widow lecturer can give the feedback of the student by looking at the student progress.

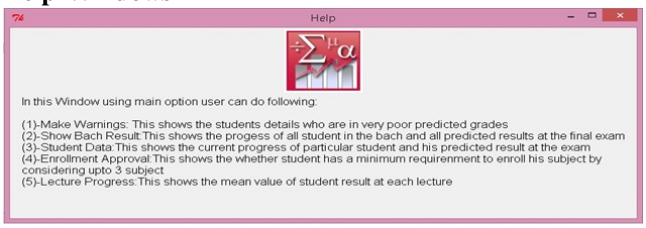


(Fig-4.18) Final predicted result of hale batch



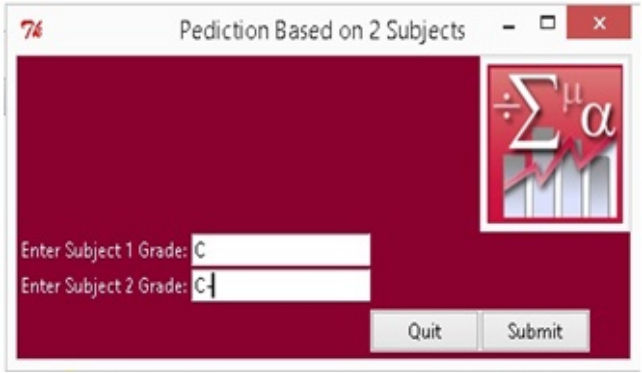
(Fig-4.19) View warnings. This window shows the students who are in F grade according to their predicted result.

Help Windows



(Fig-4.20) This window contains the directions and what are the inputs are needed.

Main functional windows



(Fig-4.21) Prediction based on Hidden Markov model input

(Fig-4.22)Prediction based on Hidden Markov model output

V. SYSTEM TESTING & ANALYSIS

The system is tested in unit testing in each and every algorithms and function. And after integrating the algorithms system is tested for accuracy testing using previous batch data and actual results and those result are displayed in graphically to make easy to identify the percentage of accuracy. Each UI is tested for appropriate function call functions like when button click using python mock framework. And final performance is tested using manually and increase

the performance by re editing the system in functions like Makcov prediction re edit the system to predicted value in 4 levels (A, B, C and F) instead of 10 levels (A+, A,A-,B+) by deciding that accuracy level I more enough for prediction. Security of the data base tested using creating users in the database with different privileges (ex lecturer only can edit the mark and student and other user could see only the marks that related to their user name)

Performance of the system tested manually and edits the system as mentioned in test approach section. Security of the system improved by using encryption method in usernames and passwords and also in database creating granted users. Failures of the system is handle by keeping the backups of a important data files on database something like configuration files to successfully face to the various or text file crashes. Lecture can store other data files in database if he wants. If some intruder hack the system or lecture forget the system password he can restore the password to the previous state by using memento design pattern.

VI. CONCLUSIONS AND FUTURE WORK

By implementing this system I intended to take more accurate prediction on students likely grade in an exam by taking prediction in both ways and get average value. For comparison with other student who has registered with that system for same course I propose to use graphical representation of performances with respect to multiple variables. By implementing second part another analysis can be to assess how each course module depends on each other, how they impact each other for the performance of the student. For this, a correlation between grades (CA and exam) can be taken. Further, predicting results of a subject based on the performance of the other subjects.

In this thesis, I develop an extension of the Hidden Markov Model (HMM) that addresses 2 most important challenges of student studying pattern: non-stationary and non-linearity. Specifically, I extend the HMM to include a novel exponentially weighted Exponential regression (ER) algorithm to handle these two challenges. I show that this extension allows the HMM algorithm to model not only sequence data but also dynamic student CA marks at several weeks. I show the update rules for the HMM parameters can be written in a form of exponential moving averages of the model variables. I further propose a double weighted ER algorithm that is able to adjust sensitivity of CA marks.

On my view this system successfully achieves the each objectives and deliverable which mention in the proposal and main requirement of the system is considerably sufficient for use. Recourse usage of the system is considerably high which gives high performance output to the user. Since system use text files to read inputs and since in UI each window has help option to give hints to how to use the system this system is improved its user friendliness. Graphical representation of system make easy to improve the understandably of system output. Since this system considering 100 into 4 student graphical representation is best way to represent output. As a future work I would like to improve the prediction accuracy by introducing more machine learning algorithm like nearest neighbor apart from the Hidden Macon model and regression. And also would like to extend system to use Linux platform standalone system as well. Currently system is reading data from text files I would like to extend this format to read input from excel sheets as lectures are commonly user excel sheets for their works

REFERENCES

[1] Yong Yuan, Spryer Netflixs Predictive Auto Scalar, 2013, Internet: http://techblog.netflix.com/2013/11/scryer-netflixs-predictive-auto-scaling.html (17 June 2015).

[2] Phil Blunso, " Hidden Markov Models" , Aug. 2004, Internet: http://digital.cs.usu.edu/ cyan/CS7960/hmm-tutorial.pdf (11 July 2015).

[3] David Harper, " Exploring The Exponentially Weighted Moving Average", Internet: http://www.investopedia.com/articles/07/ewma.asp (19 July 2015).

[4] Lahiru Sandaruwan , " Apache Stratos Auto Scaler" ,2015, Internet: http://mailarchives.apache.org/modmbox/stratosdev/201411.mbox@mail.gmail.com (25June 2015).

[5] scikitlearn.org, Regression models", Internet: http://scikitlearn.org/stable/supervisedlearning.htmlsupervisedlearning(10 Aug 2015).