Smart Analyzer: Assisting College Management through Machine Learning and Data Analysis

Seema Kedar[1], Saurabh Sutar[2], Himanshu Prasad[3]

[1]Head of Dept., Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Pune
[2]Student, Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Pune
[3]Student, Computer Engineering, JSPM's Rajarshi Shahu College of Engineering, Pune

Abstract: This work explores various opportunities to improvise regular tasks done by college faculty viz. Exam Result Analysis, Daily Student Attendance Analysis and Lecture Schedule Storage. Result Analysis becomes a tedious task when handled through traditional pen-paper methods and spreadsheets. This can be simplified by using Classification and Regression techniques. Through Regression, module-wise clarity of subjects can be foretold for students. Classification and Clustering algorithms can help to segregate students in various groups so that additional efforts can be taken for slow learners. It can also be used for classifying modules of a specific subject based on their complexities and course outcomes. The usage of register files for daily student attendance can be improved in a digital approach through Android and Django Framework. Through this approach, attendance can be tracked regularly and lecture (session) wise analysis can be done without the clutter of traditional pen-paper approach. Besides, for storing Lecture schedules and relevant timelines in the Realtime database furnishes additional benefits involving access to multiple users simultaneously. Technologies like Django Framework, Android OS, Realtime Database Systems and Machine Learning algorithms make these tasks simplified and less time-consuming. Data Analysis of Exam Results can be used for classifying student response to the teaching-learning process and can help in strategic outlining for future enhancements. Results of the proposed system consists of graphical representation of analysis done on input data and real time analysis of attendance data.

Keywords: Classification, Regression, Clustering, Django Framework, Data Analysis, Realtime database system, Android.

1. Introduction

As a traditional approach, many of the lecturers in universities and colleges use conventional methods of pen and paper to mark a student's attendance. Also, lecturers are supposed to manage the daily lecture schedule accordingly through the non-automated approach. Also, Analysis of results of internal and university exams turns into a tedious task if done manually. Our main motive is to reduce and automate these external tasks instructors need to perform and turn their entire focus on the teaching-learning process. Also, the work aims to reduce unnecessary paperwork coming under these factors. The proposed system consists of three primary modules namely Exam Result Analysis, Daily Student Attendance Analysis and Lecture Schedule Storage. For Exam Result Analysis module, proposed system provides results in the form of graphical analysis through clustering algorithms for student entries and marks scored in college and university exams. This can be helpful to learn capabilities of a student based on their performance and individual statistics of the same for the recommendation of further efforts to be taken by instructors and students. For Student Attendance Analysis module, the focus is on providing analysis features like daily/monthly statistics, class-wise, subject-wise and instructor-wise analysis, displaying subject wise and overall defaulter lists for a class, etc. Through these analysis, necessary actions on developing the teaching-learning process can be done. For Lecture Schedule Storage, process of creating and storing daily lecture schedules and modifying it according to syllabus completion and instructor availability. We have provided features like Creating and modifying of daily schedules, arranging extra slots for extracurricular activities and workshops, special features for privileged users like arranging staff meetings, declaring and managing holidays and arranging special activities exclusively for instructors and college staff. The main motive behind providing modification feature is to optimize the time utilization of students and college hours. Briefly, the proposed system for college faculty provides following features:

1. Introduce Exam Result analysis for Internal and University exams with help of Machine Learning algorithms and Data Analysis.
2. Analyze Daily Student Attendance for productive results
3. Enhance Lecture Schedule Storage.

2. System Description

A. Project Scope
The proposed system deals with environment confined to a specific college department.
B. User Classes and Characteristics
1. Instructors/Faculty Members:
Basic features like Manage Schedule, Monitor Attendance, Analyze Exam Results.
2. Head of Department and other privileged users:
Admin Provisions like maintaining the Database, Server access, Scheduling meetings and workshops and Basic Features.
C. Operating Environment
The main Objective of this project is to introduce machine learning algorithms in analysis of student exam results and daily attendance and simplifying lecture schedule storage management. The proposed system deals with environment confined to a specific college department.
D. Design and Implementation Constraints
Users are expected to be have a registered account for accessing the system. In this case, specific users can be given access to administrator privileges for maintaining the entire database.
E. Assumptions and Dependencies
User must have an Android device with OS version 5.0 or higher and a computer system with internet connectivity and modern browsers (e.g. Google Chrome, Safari, etc.) System will be deployed on a third-party server, accessible to the user base.
F. System Features
1. Django Framework:
The proposed system will partly work on Django Framework v2.2.x introducing web-based application and MVC architecture. [14]
2. Android:
Lite Features of the system can be accessed through user's Android Devices having minimum version of Android 5.0 Lollipop. For example, users can use attendance analysis module to analyze daily student attendance on their smartphones rather than desktop systems.
3. Realtime Database System and Backend AWS Database:
Proposed system will utilize a general module for storing and analyzing data on Realtime Database system (e.g. Google Firebase [13]) for prompt results and precise analysis report. The System also is integrated with Amazon S3[15] to provide reliable and secure storage of student and instructor data for future use and analysis.
4. Machine Learning Libraries:
Open Source Python ML libraries like NumPy, Scikit-learn, TensorFlow, etc. are intended to be used for implementation through Python 3.x
G. External Interface Requirements
User Interfaces:
1. Simple Interface
2. Easy Navigation
3. Strategical use of color and texture
4. User centric approach
5. Provide helpful information
Software Interfaces
1. Operating System:
Windows 7/8/10, Ubuntu 18.04+
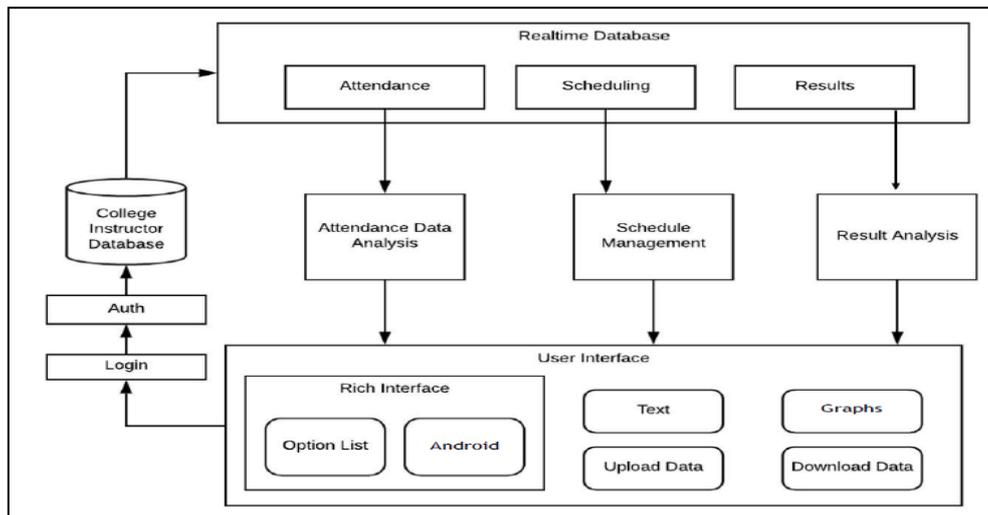2. Database:
Google Firebase Realtime, MySQL, Amazon S3.
3. Programming Languages:
Python version 3.6.x, HTML, CSS, JavaScript and supporting libraries
IDE:
Android SDK, PyCharm, MS Visual Studio

H. System Architecture

.

Figure 1. System Architecture of Smart Analyzer

Figure 1 portrays system architecture of the proposed system. It includes layered structure of modules comprising functionalities of three primary modules namely Result Analysis, Schedule Management and Attendance Data Analysis. Every primary module can be coupled with designated Realtime Database and a shared College instructor database for data related with considered students. User Interface precisely comprises of Web-app designed using Django Framework for desktop systems and Android application for remote attendance.

3. Experimental Data

[1] In this study, students' attendance and exam results data of a university, (in this paper named as "university X"), are used. The university X is recognized as private engineering university in India. The data considered consists records of 150 students admitted to the department Y of the university X in academic year 2019 - 2020. Following table explains considered parameters for the system –

Table 1. Variables used in this study

| Type	Parameters	Source
Explanatory Parameters	1. Attendance rate 2. Result rate	Attendance data Result data(.xlsx) file.

4. Numerical Experiments

A. Predictive Modeling using Machine Learning

Machine learning helps to predict and analysis the data considered for the system. Multiple approaches to achieve specific results were tried and are used for the same. For example, [3] Linear Regression and Support Vector Machines are used to predict the student performance. Also, various patterns related to the same can be analyzed with help of clustering algorithms. [9] To enhance the performance of clustering, system uses density-based algorithms like DBSCAN, to define the clusters accordingly.

B. Smart Analyzer for prediction and analysis

The purpose of this study, the subsequent analysis of number of present students and their presence in number of days are represented using linear regression. For the analysis, the dataset from the user is uploaded to the system which visualizes the data. The dataset is classified using DBSCAN clustering for classifying the students as defaulter and non-defaulter students as per attendance [4]. The attendance module has current attendance correlated to Student ID. An Android Application is provided to the admin for taking attendance of students from which the collected data will be fetched by the system from where the prediction is done. Fig. 2 represents standard flow of machine learning algorithms used for the proposed system.
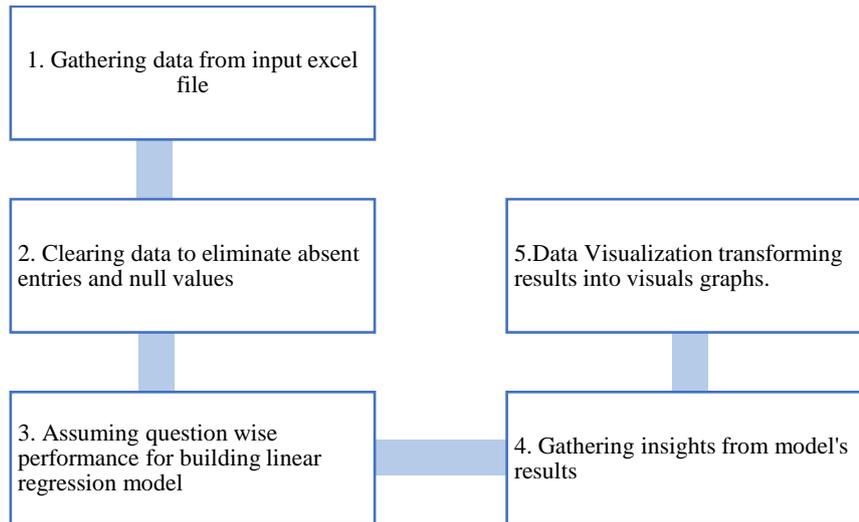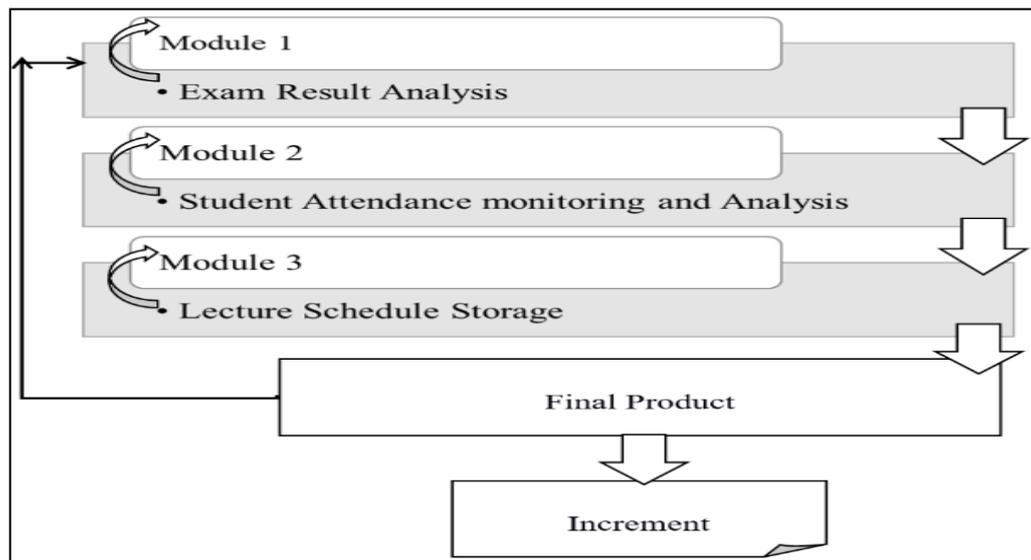
Figure 2. Machine Learning process for Smart Analyzer

4. System Implementation

The system is divided into 3 distinct interrelated modules:
1. Exam Result Analysis
2. Student Attendance Analysis
3. Lecture Schedule Storage
Following Fig. 3 shows module wise implementation plan of the proposed system considering the above primary modules

Figure 3. Module-wise implementation plan

Module 1: Exam Result Analysis -
In this module we have focused on analyzing student exam results. This module is further divided into following fragments:
1. Functionalities for uploading results in specified spreadsheet format.
2. Displaying graphical analysis.
3. Displaying Categorization of students based on results and recommending modification in schedules.

.

Fig. 4.1 and 4.2 shows results for 'Question-wise' analysis of score of the students in considered exam result dataset. [8] Linear regression performs the task to predict the number of students represented on the Y-axis based on a given independent variable (x) i.e. the number of questions in the dataset. In the visualization output of Fig. 4.1, X (input) defines questions according to their serial number and Y(output) is the number of students attempting that questions. Fig. 4.2 shows linear regression output for each question based on the scores of students for that question.

Hypothesis function for Linear Regression: $y = \theta1 + \theta2.x$
While training the model we are given:
x: input training data (univariate), y: labels to data (supervised learning)
θ_1: intercept, θ_2: coefficient of x
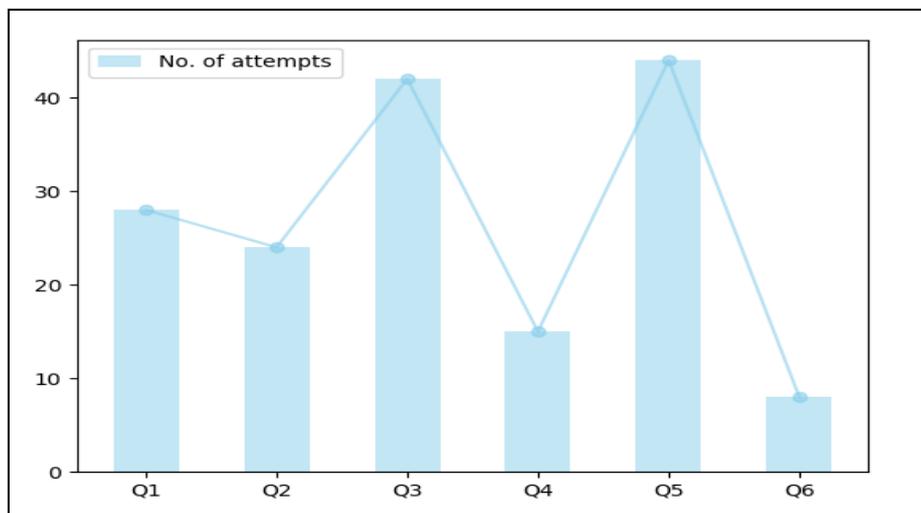
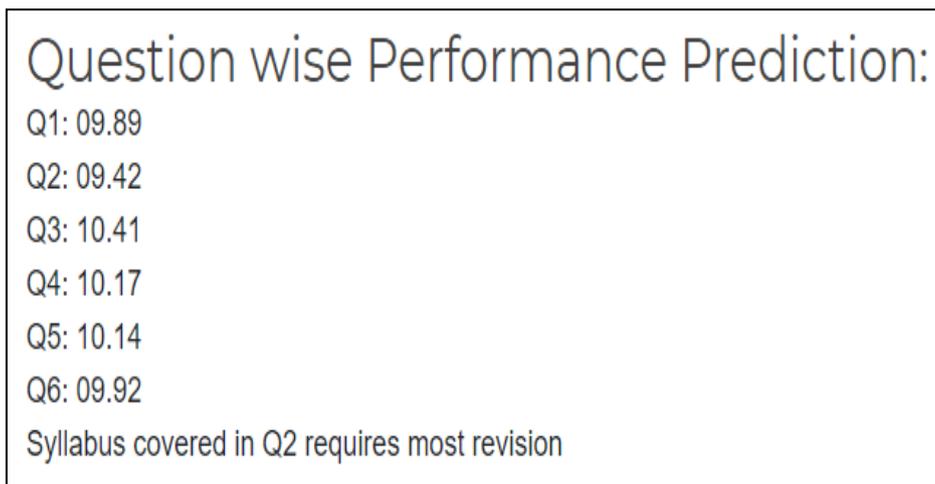**Figure 4.1.** Question-wise Preferences

Figure 4.2. Question-wise Performance prediction through Linear Regression
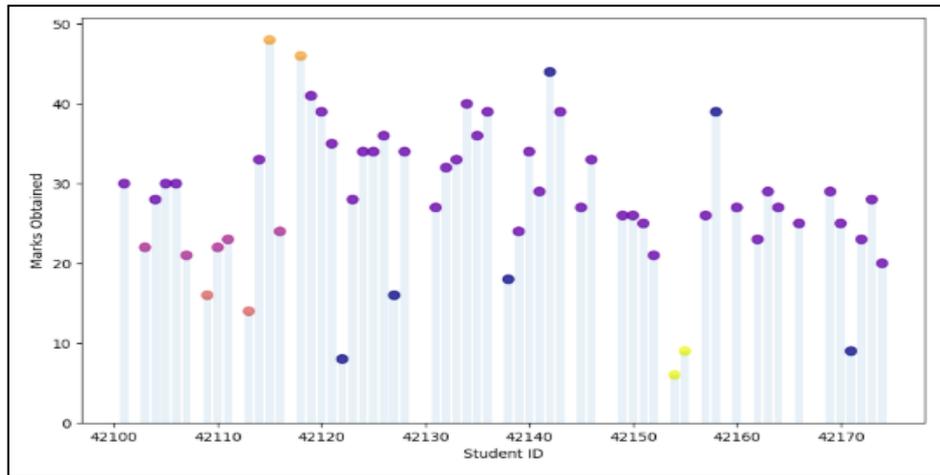
Figure 5. Student performance analysis using DBSCAN

The DBSCAN algorithm is based on this intuitive notion of "clusters" and "noise".[9] In this module, DBSCAN algorithm is utilized to perform clustering of student data according to the density of the total marks obtained by a particular student. In this fashion, number of students can be categorized based on their performance in an exam. Students can be basically categorized as "Advanced Learners" and "Slow Learners" which can be helpful for users (teachers and instructors) to maneuver their focus for these students separately for that particular subject.

Figure 6. Output for Exam Result Analysis

Module 2: Student Attendance Monitoring and Analysis—
In this module we will focus on developing student attendance monitoring system. This module is further divided into following fragments:
1. Setting up Realtime Database
2. Functionalities for taking attendance through Android Phones
3. Functionalities for Attendance Analysis
4. Functions for privileged users
This module involves techniques like Android, database management and specific algorithms for analyzing the attendance.
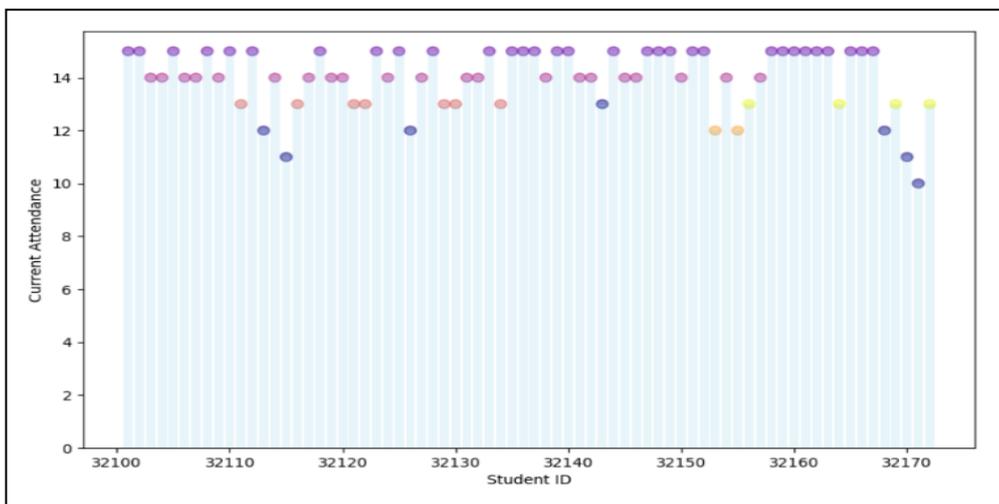
.

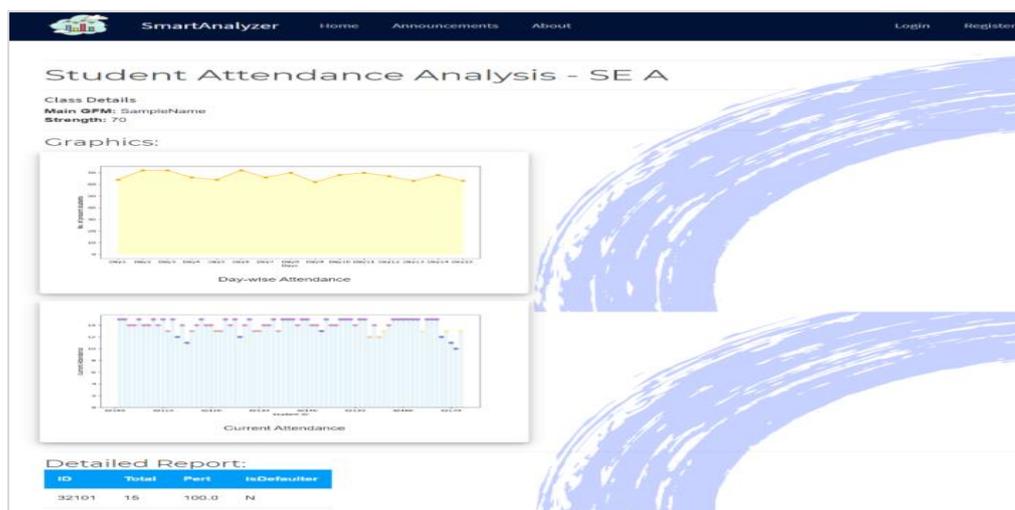
Figure 7. Attendance Analysis Using DBSCAN.

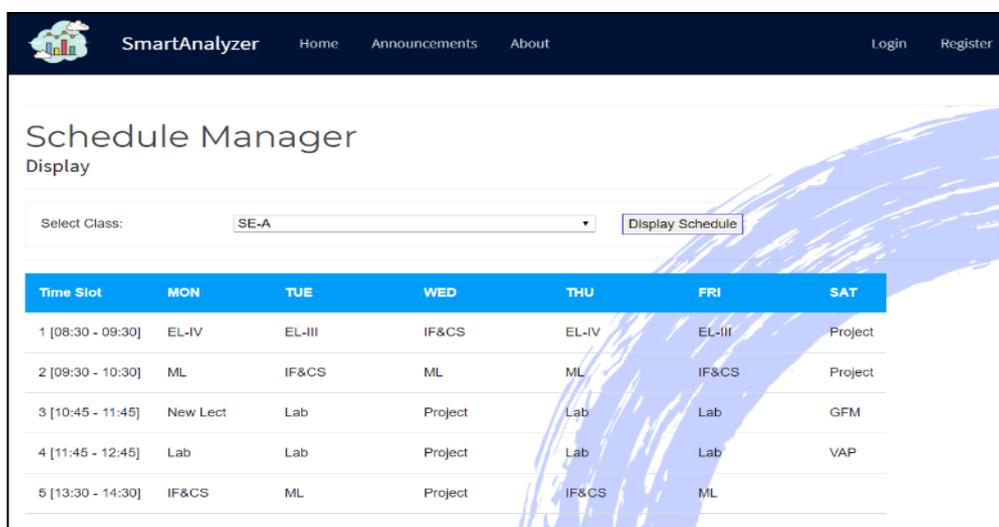**Figure 8.** Output for Student Attendance Analysis

Figure 8. Output of Schedule Manager

*Module 3: Lecture Schedule Storage –*

In this module we will focus on developing daily schedule management system for instructors and Senior authorities. This module is further divided into following fragments:

1. Setting up Realtime Database

2. Functionalities for Creating and Modifying the schedule e.g. functionality for adding a slot for Extra-curricular activities and workshops.
3. Functionality for setting up a staff meeting for privileged users.
This module involves techniques like Realtime database updating and declaration and database management.
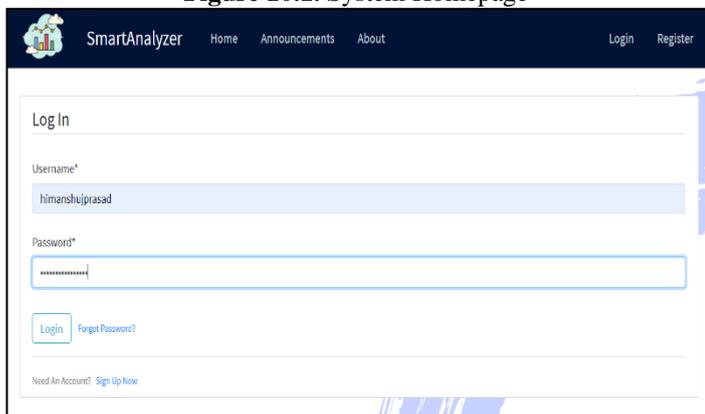
5. System UI

Following are some screenshots of developed web application in Django Framework:
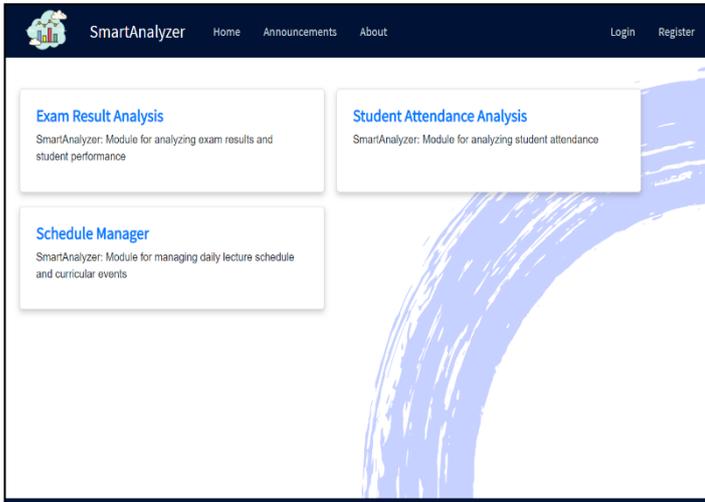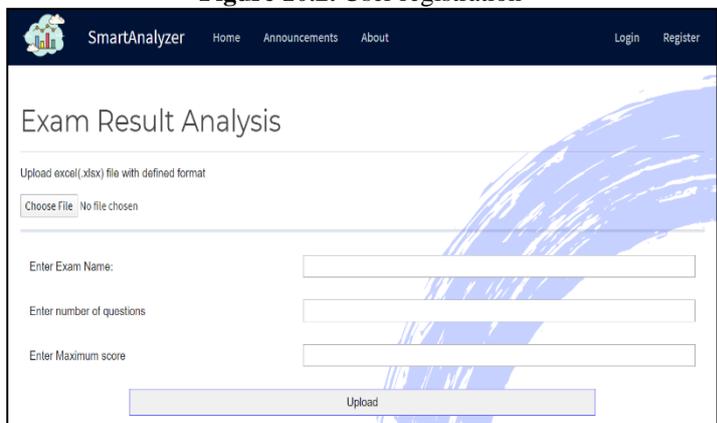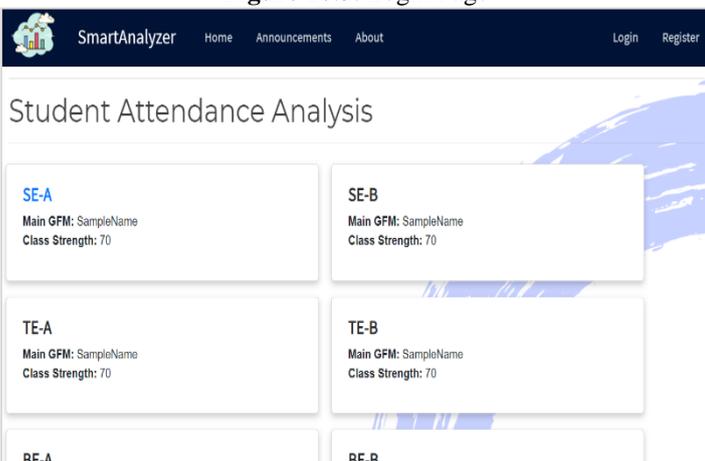

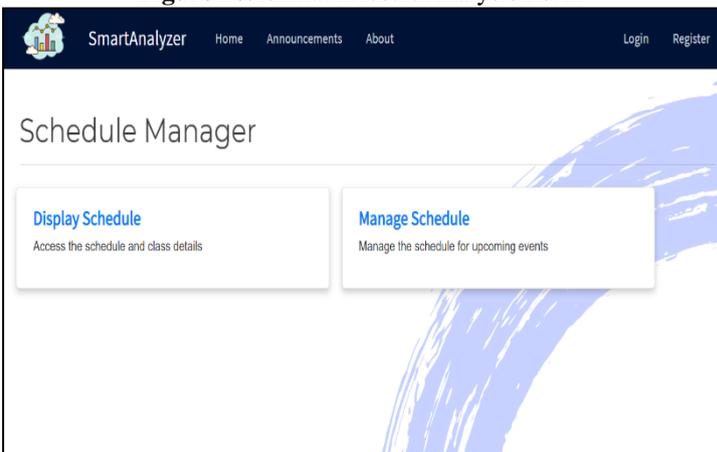**Figure 10.1.** System Homepage
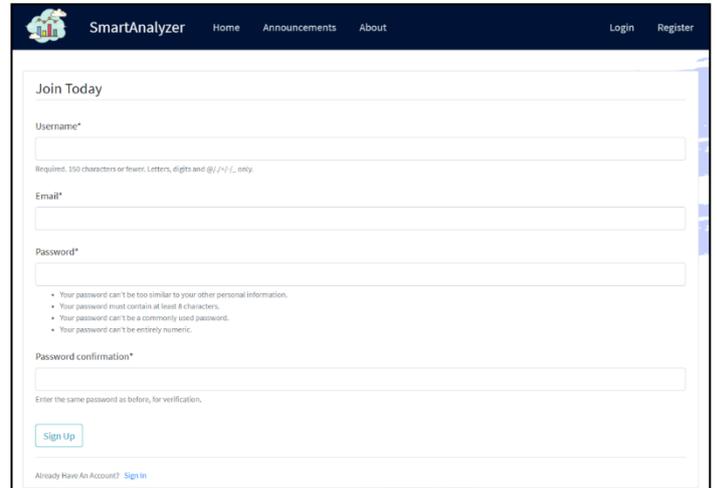
Figure 10.2. User registration

Figure 10.3. Login Page

Figure 10.4. Exam Result Analysis Form

Figure 10.5. Student Attendance Analysis Page

Figure 10.6. Schedule Manager Page

.

7. Conclusion and Outlook

Hence, we can observe that traditional approaches for taking class attendance, managing daily lecture schedule and analyzing exam results involve lots of paperwork and unnecessary tasks which can be optimized through digital platforms and technologies viz. Machine learning and Data Science approaches through algorithms like clustering, linear and logistic regression, decision trees and setting up global Realtime database for each module. Additionally, proper scheduling and result analysis can help and enhance teaching-learning process and student results.

8. Future scope

In this research work, some considerations with respect to the modeling of system in Smart Analyzer are made, such as introduction of RPA to handle automated repeatable tasks which required Faculty as admin to perform. These tasks can include queries, processing and maintenance of database. Moreover, the system can analyze PDF format documents too. The web application can be integrated with cloud storage viz. Amazon S3, etc. additionally to provide enhanced overall security [15] and store student database for future reference.

References

1. Monica Ciolacu, Ali Fallah Tehrani, Rick Beer and Heribert Popp, "Education 4.0 – Fostering Student's Performance with Machine Learning Methods", IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), 29 Oct 2017.
2. Rachel Baker, Brent Evans, Qiujie Li, Bianca Cung, "Does Inducing Students to Schedule Lecture Watching in Online Classes Improve Their Academic Performance? An Experimental Analysis of a Time Management Intervention", Springer Nature B.V. 2018, 28 July 2018.
3. Raza Hasan, Sellappan Palaniappan, Abdul Rafiez Abdul Raziff, Salman Mahmood and Kamal Uddin Sarker, "Student Academic Performance Prediction by using Decision Tree Algorithm", IEEE 2018, 4th International Conference on Computer and Information Sciences (ICCOINS)
4. Nobuhiko Kondo, Midori Okubo, Toshiharu Hatanaka, "Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data", 2017 IEEE 6th IIAI International Congress on Advanced Applied Informatics
5. Fahad Razaque, Nareena Soomro, Shoaib Ahmed Shaikh, "Using Naïve Bayes Algorithm to Students' bachelor Academic Performances Analysis", © Springer International Publishing AG 2017
6. Zun Hlaing Moe, Thida San, Hlaing May Tin, Nan Yu Hlaing, and Mie Mie Tin , "Evaluation for Teacher's Ability and Forecasting Student's Career Based on Big Data" Springer Nature Singapore Pte Ltd. 2019
7. Leonard K.M. Poon, Siu-Cheung Kong, Thomas S.H. Yau, Michael Wong, and Man Ho Ling, "Learning Analytics for Monitoring Students Participation Online: Visualizing Navigational Patterns on Learning Management System" Springer International Publishing AG 2017.
8. Amand F Schmidt, Chris Finan, "Linear Regression and Normality assumption", 2017 Elsevier Inc.
9. Xiao Xua, Shifei Dinga,b, Zhongzhi Shib "An improved density peaks clustering algorithm with fast finding cluster centres" , Elsevier, May 2018.
10. Leonard K.M. Poon, Siu-Cheung Kong, Thomas S.H. Yau, Michael Wong, and Man Ho Ling "Learning Analytics for Monitoring Students Participation Online", Springer 2017.
11. Zun Hlaing Moe(&), Thida San, Hlaing May Tin, Nan Yu Hlaing, and Mie Mie Tin "Evaluation for Teacher's Ability and Forecasting Student's Career Based on Big Data" Springer, 2019.
12. "Linear Regression – Detailed View" https://towardsdatascience.com/linearregression-detailed-view-ea73175f6e86
13. "Realtime Database and Applications of Realtime Database" https://en.wikipedia.org/wiki/Real-timedatabase , https://firebase.google.com/
14. "The Django Project", https://www.djangoproject.com/
15. "Amazon S3 – AWS Documentation", https://docs.aws.amazon.com/s3/index.html