Course Scheduling Using Genetic Algorithms Enhanced by Linear Regression for Data Mining Course Participants

Arie Susetio Utami[1*], Agust Isa Martinus[2], Freddy Wicaksono[3], Rangga Manggala Yudha[4]

[1,2,3,4]*Faculty of Engineering, Muhammadiyah University of Cirebon, Indonesia*

*corr-author: arie.utami@umc.ac.id

Abstract – Course scheduling at the beginning of each semester is an absolute must, considering changes in course instructors, changes in the availability of lecture schedules, changes in lecture infrastructure in terms of number, capacity, and time of use, changes in the number of lecture participants, both new and repeat participants. This research aims to design an optimal scheduling system by considering scheduling constraints to avoid conflicts and increase the effectiveness of lecture scheduling management. The linear regression method is used to predict the number of lecture participants using the 2019-2022 academic year data as model data and the 2023 academic year data as testing data to validate the prediction data. Lecture scheduling uses a Genetic Algorithm with a fitness function for the number of cross-class schedules - contracted by repeating students - that conflict with the chromosomes used by courses, classes, lecturers, rooms, schedules, and others. The designed scheduling system has a prediction model with high accuracy and a coefficient of determination (R-Sq.) above 95% and RMSE below 10. This scheduling system is efficient, minimizing scheduling conflicts to 0 percent.

Keywords: course scheduling; Genetic Algorithm; Linear Regression; course S; prediction

I. INTRODUCTION

Course scheduling is an essential activity that necessitates focus for the learning process to be efficient and effective. Scheduling issues frequently challenge higher education administrators, particularly within academic programs [1]. Many resources are considered while preparing lecture schedules, such as the number of students, classes, lecturers, courses, rooms, and lecture times [2].

The number of course participants must be estimated based on the number of students enrolling. The number of course participants is defined by the decay of the number of students in each class and the number of students who repeat the course. In scheduling, a mismatch often occurs between the number of students and the number of courses opened, resulting in every course contract procedure opening and closing class because the number of participant quotas is whole or too small. The number of [3] classes required each semester must be organized per resource allocation and room utility to avoid conflicts between courses, rooms, lecturers, and lecture hours. Scheduling will be a routine task with the same problem every semester if the academic information system cannot satisfy all these requirements [4].

Based on these issues, it is necessary to design an optimal scheduling system by considering scheduling constraints and predicting the number of participants to minimize or avoid scheduling conflicts and increase the administrative efficiency and effectiveness of reliable lecture scheduling management. This scheduling design is expected to provide solutions for study program managers by optimizing the utility of resources, classrooms, time, and other facilities to improve the quality of educational services to students, which will positively impact the student learning experience [1,2].

Linear regression is used to predict the value of a response variable based on regressor variables [5]. Several linear regression studies were used to predict the number of college participants in higher education institutions [6,7]. The research shows the accuracy of regression analysis in predicting college participants in certain courses based on historical data, namely student graduation in the course in the previous semester [8,9]. However, the regression model built to predict the number of participants has yet to explain the factors that affect student graduation because, practically, there is more than one possible factor to influence the number of course participants and because each study program also produces different regression models. The number of classes that will be opened in the new semester can be determined based on the results of predicting the number of course participants. The number of classes will be the input of one of the chromosomes in scheduling lectures using genetic algorithms [8].

Effective time management and scheduling are crucial skills for individuals and businesses in the contemporary era. Implementing an efficient scheduling system has been demonstrated to result in a reduction in time expenditure, an increase in production levels and a concomitant decrease in stress levels [10,11]. It is, therefore, imperative that effective scheduling tools are created. Genetic algorithms represent one of the most promising approaches to resolving scheduling problems [2,4]. Genetic algorithms are a class of search and optimisation techniques inspired by the principles of natural selection [12]. The technique employs natural selection, mating, and mutation to identify optimal solutions to complex problems without a clear mathematical model. This approach enables the generation of the most effective schedule, considering resource availability, time constraints, and job priorities. Implementing a genetic algorithm has been demonstrated in the context of lecture scheduling. Genetic algorithms have been widely utilised to determine optimal scheduling. In recent research, genetic algorithms have been shown to produce optimal lecture scheduling by integrating regression models and student time preference parameters with class availability [12].

The integration of linear regression with a genetic algorithm enables the accurate prediction of the number of students, thus facilitating the utilization of data on student numbers as input for lecture scheduling parameters to minimize scheduling conflicts [12,13]. The specific data context is applied to a case study at the Engineering Faculty of the Muhammadiyah University of Cirebon, enabling the development of optimal and relevant prediction and scheduling models that align with the faculty's needs. Furthermore, this approach can be extended to propose a general model for implementation in higher education. Therefore, this research aims to design and develop an optimal course scheduling system by integrating linear regression and genetic algorithms. The linear regression model accurately predicts the number of students for each course, while the genetic algorithm generates a conflict-free and efficient course schedule. This integrated approach is expected to minimize scheduling conflicts, enhance the efficiency of academic administration, and improve the overall quality of educational services.

II. METHOD

A. Data Collection

The data was gathered from the academic records of students enrolled in the institution over five years, comprising both odd and even academic years between 2018 and 2022. The data was downloaded from the university information system in the form of study result cards. The data set 2024 in Table 1 comprises the following variables: student identification number (NIM), course, grade, semester, year, faculty, study program, and quality value (Value). Information on time attributes and lecturers was obtained from the faculty.

B. The Processing and Validation of Datasets

The collected datasets (Table I) are then processed and validated to achieve data consistency for later analysis. The aim is to produce the required information following the format of the predictive model. The collected data is observed and analysed by converting the data set into a structured data form according to data mining needs; this process is called data set pre-processing [2]:

1) *Format Modification*: Change the format of the report card to the tabulated form of an Excel file (.csv) and change the format of PDF grades from the academic system to an Excel file (.csv).

2) *Clean-Up Missing Data*: This process removes files with unreadable data from the PDF to the CSV format. This process is carried out due to the large variation in the location of the data cells read by CSV, so that it reads the conversion results from the PDF.

3) *Attribute Selection*: Selection of attributes to be displayed. Some attributes do not need to be displayed. Then, the attribute will be deleted so that only the attributes of the academic year (Year), the name of the course, the number of lecturers, the number of students who passed the course and the average value of the course based on the quality letter will be deleted attributes such as lecturer name and class. The added attributes are year, semester and study programme.

4) *Attribute Grouping*: Grouping the dataset in Table II based on the attributes of academic year 2022, all semesters and Informatics study programme; this is done because the data will be regression-modelled per study programme.

5) *Character Uniformity*: Uniformity of the same course name but written differently; this is done because several courses are the same but have different names (Fig. 1) [8].

6) *Course Selection* [8]: Elective courses are removed from the list at this stage, leaving only compulsory courses. Additionally, old courses that are no longer in the curriculum and new courses that are in the new curriculum are deleted to ensure that the

predicted number of course participants is equivalent to that of past data courses in the old curriculum and courses in the new curriculum (see Table III).

C. Regression Linear Model

In order to predict the number of course participants, a linear regression model is employed to describe the relationship between two variables, namely the influencing or independent variables (X) and the dependent variable (Y) [6,14]. This regression model is utilized because it can estimate or predict new data based on time series data from the past. The form of the multivariate regression analysis model is expressed in (1)

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (1)$$

Notation,

Y : Number of students in each course
x_1 : Number of passing students
x_2 : average course grade
b_1 : coefficient for x_1
b_2 : coefficient for x_2

Regression equation matrix shows in (2) [5]:

$$\hat{y} = xb \quad (2)$$

where (3),

$$\hat{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, x = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (3)$$

Regression coefficient matrix (4) [5],

$$b = (x'x)^{-1}x'y \quad (4)$$

Statistical inference on the regression model regarding the significance of the regression model [5,6]:

- Analysis of Variance (regression model significance test).
- T-test (Test of the significance of regression coefficients).

The linear regression model constructed using (1) from the preprocessed dataset is divided into two parts (see Fig. 4), namely:

- The data model uses data for the academic years 2019 to 2022. The dependent variable (y) is the number of course participants. There are two independent variables determined, namely the number of students who pass the course (passed) as independent variable one (x_1) and the average value of the course based on the quality score as (x_2). The model data is used to determine the regression model. The best regression model will be determined by the stepwise method with the criteria of the R-Sq value or the coefficient of determination close to 1 or 100%.
- Testing data uses the number of course participants in the 2023 academic year. The best regression model obtained from the Model Data is then used to determine the predicted value using the 2023 academic year data.

D. Validation of Regression Model Prediction Results

The best regression model obtained from the model data is then applied to the data on the number of college participants in the 2023 academic year to predict the number of college participants in the next academic year. After obtaining the predicted number of participants in 2024, the next step is to calculate the prediction error value by determining the difference between the actual number of college participants in 2023 and the predicted value of the number of college participants. Root Mean Square Error (RMSE) is the validation method or predicted error. Prediction results are called good if the RMSE value is close to zero [15,16]. The RMSE value will be aligned with the R-Sq value or the coefficient of determination, where the R-Sq value close to 1 or 100% will provide information on prediction results with a small error or error against the actual value (predicted value) [9,17]. Then, the validated prediction results become the input of one of the chromosomes in scheduling using genetic algorithms as in (5) [6,7].

$$RSME = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \quad (5)$$

Where,

y_i : actual data/testing data 2023
$\hat{y}_i$: model Model prediction result data
i : dataset order
n : number of data

Fig. 1 Example of course name uniformity

TABLE I
DATASET 2024

No	NIM	Course	Grade	Semester	Year	Faculty	Programme	Value
1	220511047	Religion	A	Odd	2022	Engineering	Informatics	4
2	220511171	Religion	C	Odd	2022	Engineering	Informatics	2
3	220511164	Agama	A	Odd	2022	Engineering	Informatics	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21695	230411106	Probability	B	Even	2023	Engineering	Industrial	3
21696	230411033	Probability	A	Even	2023	Engineering	Industrial	4
21697	230411011	Probability	B	Even	2023	Engineering	Industrial	3

TABLE II
ATTRIBUTE GROUPING

Course	Sum of Participants	Sum of Passed	Sum of Value
Religion	139	138	3.65
Al Islam Humanity and Belief	143	143	3.99
Al Islam Worship, Morals & Muamalah	123	123	3.58
Al Islam and Muhammadiyah	119	113	3.82
Islam and Science	53	53	3.49
Object Oriented (Analysis & Design)	110	110	3.76
Algorithmic Analysis & Strategy	112	112	3.03
Computer Architecture & Organization	93	93	3.96
Indonesian (Bahasa)	134	134	3.49
English	134	134	3.49
…	…	…	…

TABEL III
NUMBER OF COURSES SELECTED

Programme	Initial	After Selection
Animal Husbandry	59	51
Industrial Engineering	70	63
Informatics Engineering	57	52

E. Course Scheduling Optimisation Using Genetic Algorithm Method

Genetic algorithms are search and optimization techniques inspired by natural selection. In the scheduling context, this algorithm Fig. 2-3, can generate optimal schedules by considering various factors such as subjects, classes, rooms, and available time slots [2].

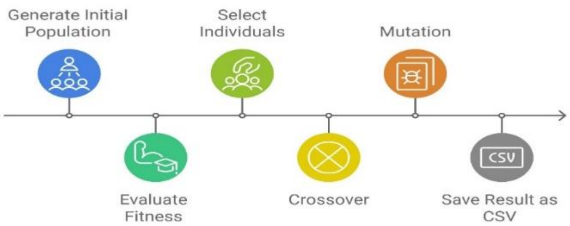

Fig. 2 Genetic algorithm process

1) Data Preparation: Before starting the genetic algorithm process, the first step is to prepare the required data, including:

- Subjects Class: The subject to be taught and the associated class.
- Room: Available classrooms.
- Available Time Slots: Available time slots for each subject and class.

This data is stored in CSV format and prepared for use in the genetic algorithm process.

2) Population Generation: The first stage in the genetic algorithm is the initial population. A set of solutions (schedules) is randomly generated at this stage. Each population member represents a different schedule, allocating subjects to classes, rooms, and available time slots [3] as shown at Fig. 4:

- Subjects, classes, rooms, and time slots are taken from the prepared data.
- Each individual is scheduled randomly to create diverse initial solutions.

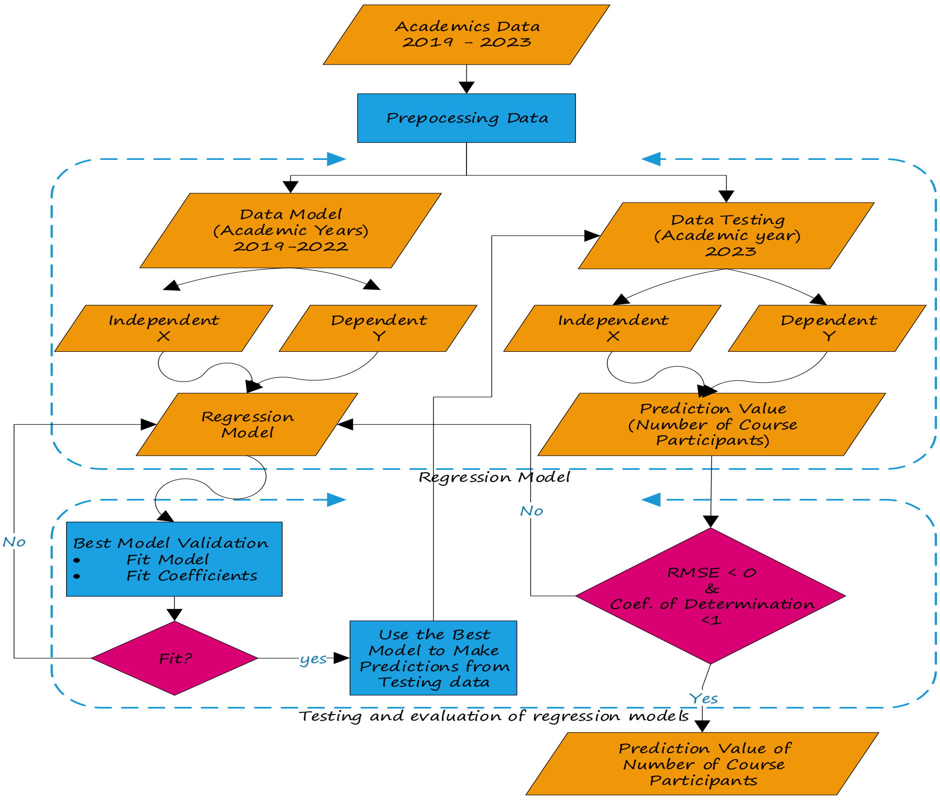

Fig. 3 Regression modelling for prediction number of course participants

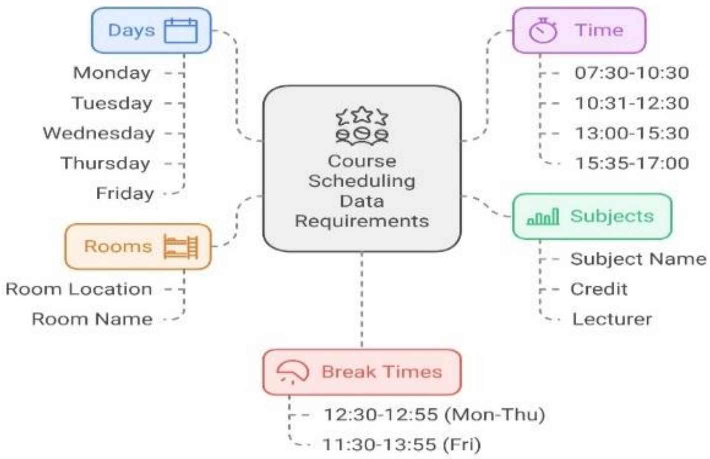

Fig. 4 Scheduling data generation

3) Fitness Function: Each individual in the population is evaluated using a fitness function. This fitness function determines how well the resulting solution solves the scheduling problem [2]. Some criteria that can be used in the fitness function include:

- No clashes between subjects taught simultaneously in the same room.
- Classroom availability is by class capacity.
- Each class gets a time slot that does not clash.
- The fewer conflicts a schedule has, the higher the fitness value.

4) Selection: Selection is the process of selecting individuals from the current population to produce offspring for the next generation. Usually, individuals with higher fitness values have a greater chance of being selected, as they tend to provide better solutions. The selection method used is Tournament Selection: A group of individuals is randomly selected, and the best of the group is chosen as parents [12].

5) Crossover: Once the individuals are selected, the crossover or recombination process is performed. Crossover is a genetic operator that combines genetic information from two parent solutions to produce new offspring. The aim is to create variation in the latest population to find a better solution [12]. Crossover in course scheduling:

- One parent provides course allocation for half of the class, while the other parent provides allocation for the other half.
- Time and space slots are exchanged between parents to generate a new schedule.

6) Mutation: Mutation is a process that introduces random changes to individual genes with a small probability. In the scheduling context, mutation can be a random time slot, space, or subject replacement change for a class. The mutation is important to maintain diversity in the population and prevent the solution from getting stuck at a locally non-optimal solution [1,18] as shown at Fig. 5:

- Randomly changing the time slot for a subject.
- Changing the space used for a particular class.

7) Saving Results: The schedule from the mutation process is then saved in CSV file format. Fig. 6 shows the framework of this study, regression as data mining for the number of lecture participants to enhance the scheduling optimization with genetic algorithms.

III. RESULT AND DISCUSSION

A. Linear Regression Model

The dataset processing results in Table IV were obtained from three study programmes in the Faculty of Engineering: Animal Husbandry, Industrial Engineering and Informatics Engineering. Model data, namely academic data from 2019 to 2022, is used to form regression models. Regression models were formed based on the academic data of each programme. The first is an initial model with all independent variables and a constant value as an intercept. The second model is the result of selecting the best model using the stepwise method; this method only includes variables that affect the regression model to predict the number of college participants according to the actual level value of 0.05. This second model is the best; this is indicated by an increase in the coefficient of determination (R-Sq) from the first model to the second model, which can be concluded that the second model of each program is able to explain data variability better than the first model so that the second model in the prediction step will be used as a model to predict the number of college participants.

B. Prediction

The regression model determined in the previous step is then applied to predict the number of college participants using testing data, namely, 2023 academic data for the three programs at the Faculty of Engineering. The prediction results using testing data obtained the results of the analysis for error or prediction error of the number of students with a small RMSE value balanced with the coefficient of determination (R-sq), which is close to 100%, which indicates that the value of the variability of the prediction error is minimum so that the prediction results have a minimum difference with the actual value of 2023 academic data (see Table V). Based on the prediction results of the number of lecturers, how many classes will be opened as one of the chromosome inputs for scheduling purposes with genetic algorithms can also be determined.

C. Initial Course Scheduling Optimisation

Scheduling lectures with genetic algorithms is done with the help of the Python programming language by integrating all academic data databases used for scheduling, as shown in Fig. 4, and the algorithm model carried out, as shown in Fig. 3. This model then produces an initial lecture schedule in the series of Fig. 6. Initial scheduling still produces lecture conflicts, with a course conflict of four courses and a classroom conflict of ten courses, with a fitness function of 22 % with conflict visualization, as shown in Fig. 7-9.

D. Evaluation of Course Scheduling Model (Improve Genetics Algorithm)

Scheduling conflicts are known based on class and time constraints and lecturer and time constraints because the results achieved in the previous genetic algorithm optimisation process still have conflicts, so based on the evaluation of scheduling performance carried out with the faculty, development is carried out by improving scheduling conflicts to minimise or even eliminate these conflicts. During the scheduling process, some time slots may not be allocated. The next step is to compare the resulting schedule with the available time slot data. Furthermore, the scheduling model in Fig. 4 is developed by adding the last four steps, as shown in Fig. 10.

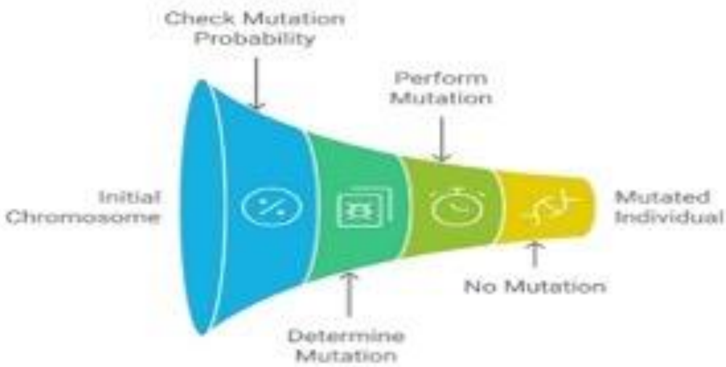

Fig. 5 Mutation process

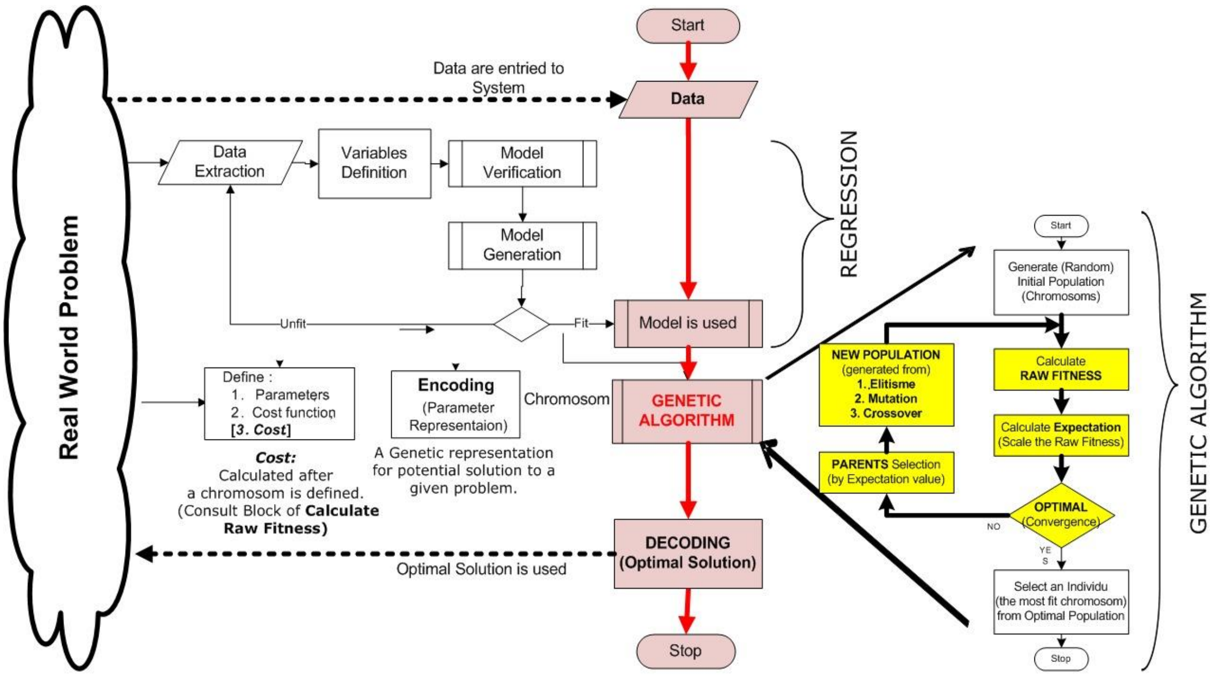

Fig. 6 GA enhanced by linear regression for data mining course participants.

TABLE IV
MODELS RECAP FROM EACH PROGRAMME

Programme		Regression Model	Anova (F value; P-value)	R-sq	Best Model
Animal Husbandry	Model 1	$Participants = 1.66 + 1.1023\ Passed - 0.5\ Ave.Grade.$	145.13; 0.000	72,16%	Model 2
	Model 2	$Participants = 1.1036\ Passed.$	2372.60; 0.000	95.42%	
Industrial Eng.	Model 1	$Participants = -0.73 + 1.0166\ Passed + 1.92\ Ave.Grade.$	1287,40; 0.000	95.69%	Model 2
	Model 2	$Participants = 1.0158\ Passed + 1.706\ Ave.Grade.$	8996.56; 0.000	99.35%	
Informatics Eng.	Model 1	$Participants = 19.41 + 1.0416\ Passed - 5.43\ Ave.Grade.$	2426.38; 0.000	97.69%	Model 2
	Model 2	$Participants = 1.05332\ Passed.$	35664.96; 0.000	99.67%	

TABLE V
RMSE AND COEFFICIENT OF DETERMINATION

Programme	RMSE	R-sq
Animal Husbandry	3.095518	95.42%
Industrial Engineering	5.017261	99.35%
Informatics Engineering	9.123243	99.67%

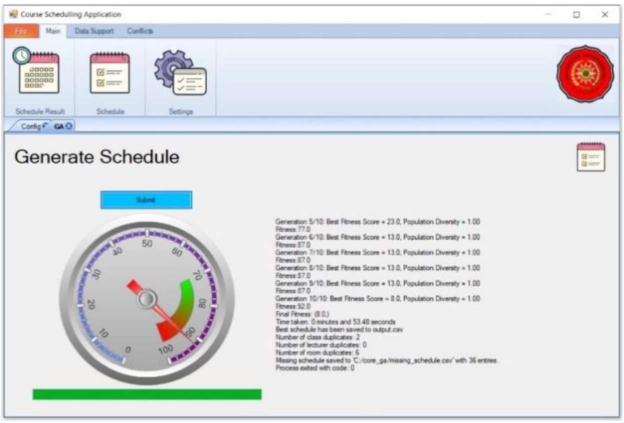

Fig. 7 Initial course scheduling

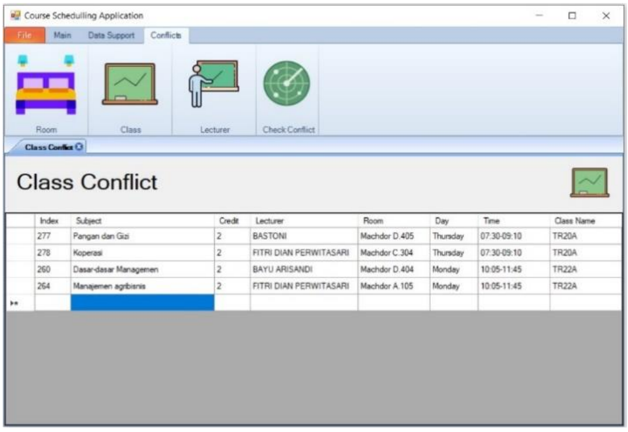

Fig. 8 Course conflict

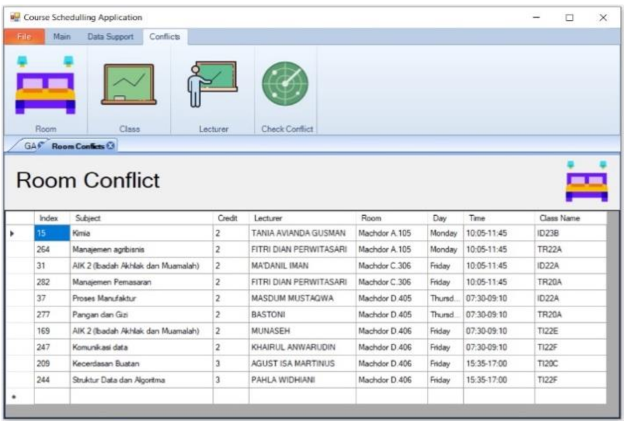

Fig. 9 Classroom conflict

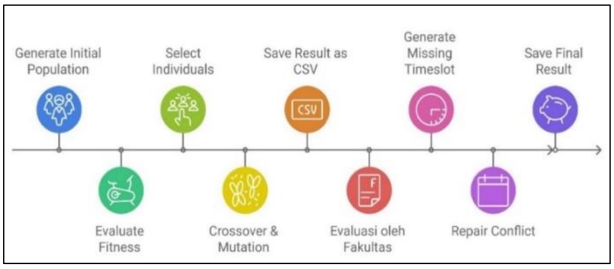

Fig. 10 Improved genetic algoritma

After the Faculty have evaluated the initial course schedule, the next step is to generate the missing time slots. From the comparison, the missing time slots can be identified. The process compares the schedule results with the available time slots, and then the unused time slots will be saved. The next step is to Repair Conflict; after the initial schedule is generated, there are often conflicts, such as two classes being scheduled in the same room at the same time. The algorithm will check the file to find clashes in the schedule. The Conflict is then fixed using the unused time slots saved in the Generate missing time slots step. The process is to detect clashes in the schedule, select alternative time slots and fix the schedule to minimise the clashes. The result is saved into a new CSV file once all conflicts are fixed and the optimal schedule is generated. This file contains the final schedule with minimal conflicts Fig. 11, ready to be used as the official schedule by the Faculty.

After improvement, an optimal schedule is produced by considering various constraints and ensuring clashes are minimized, or zero conflict is achieved in scheduling. This algorithm is very effective in situations involving many variables and constraints, such as academic scheduling. By considering the predetermined data requirements, this scheduling application will provide an optimal course schedule based on the actual conditions in the field. The results of the scheduling performance evaluation in Table VI provide reasonable predictions of the number of lecture participants with a coefficient of determination above 95% and a small RMSE value below 10 for the three study programs. The scheduling efficiency value or fitness function can be minimized to 0% or zero conflict, with a computation time of 0.8 seconds for predicting the number of students and 30 seconds for scheduling lectures using genetic algorithms.

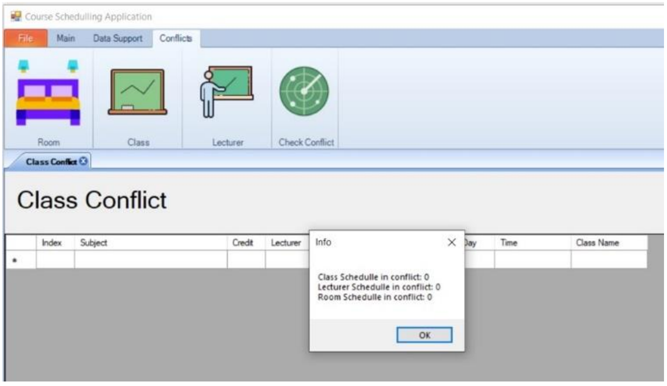

Fig 11. Improved AG: Zero Conflict

TABLE VI
PERFORMANCE OF COURSE PARTICIPANT PREDICTION AND LECTURE SCHEDULING

Metric	Linier Regression	Genetic Algorithm
Prediction accuracy	R-S > 95% & RMSE < 10	-
Scheduling efficiency	-	0% (Zero Conflict)
Computational time	0.8 sec	30 sec

IV. CONCLUSION

A scheduling system using a genetic algorithm involving a linear regression model to predict the number of lecture participants was successfully designed. The prediction model generated from linear regression shows high prediction accuracy with a coefficient of determination (R-sq) above 95% and a small RMSE value below 10 for all three study programmes. The resulting scheduling system is also efficient and can minimise scheduling conflicts by 0%. This model provides an effective solution to scheduling problems in higher education, particularly for the case study at the Faculty of Engineering, Universitas Muhammadiyah Cirebon, and has good potential to be applied in other higher education institutions.

ACKNOWLEDGEMENT

We extend our profound gratitude for their invaluable contribution to the advancement of this research endeavour. The financial support provided by the Directorate of Research, Technology and Community Service at the Indonesian Ministry of Education, Culture, Research and Technology has facilitated the execution of this research study.

REFERENCES

[1] H. Alghamdi, T. Alsubait, H. Alhakami, and A. Baz, "A Review of Optimization Algorithms for University Timetable Scheduling," *Technology & Applied Science Research*, vol. 10, no. 6, pp. 6410–6417, 2020, [Online]. Available: www.etasr.com

[2] F. Wicaksono and B. P. Putra, "Course Scheduling Information System Using Genetic Algorithms," *ITEJ (Information Technology Engineering Journals)*, vol. 6, no. 1, pp. 35–45, 2021, Accessed: Mar. 30, 2024. [Online]. Available: https://syekhnurjati.ac.id/journal/index.php/itej/article/view/55

[3] R. A. T. Sudalyo, M. B. Hartono, and F. Wicaksono, "Implementation of the Constraint Satisfaction Problems Method in Genetic Algorithms for Course Scheduling Systems," *Jurnal Improsci*, vol. 1, no. 3, pp. 140–147, 2023, Accessed: Mar. 30, 2024. [Online]. Available: https://annpublisher.org/ojs/index.php/improsci/article/view/149

[4] I. D. Wahyono *et al.*, "Optimizing Regression on Genetic Algorithms for Effective scheduling based on the Predicted Number of Students," in *Proceedings - International Conference on Education and Technology, ICET*, Institute of Electrical and Electronics Engineers, 2022, pp. 240–244. doi: 10.1109/ICET56879.2022.9990626.

[5] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientist*, Ninth. Prentice Hall, Pearson, 2012.

[6] D. C. . Montgomery and G. C. . Runger, *Applied statistics and probability for engineers*, Third. New York: John Wiley & Sons, Inc., 2003.

[7] M. L. Berenson, D. M. Levine, and T. C. Krehbiel, *Basic Business Statistics: Concepts and Applications*, Twelfth. New Jersey: Prentice Hall, Pearson, 2012.

[8] E. Ardhia Utami, Y. Sholva, A. Perwitasari, and J. H. Hadari Nawawi, "Sistem Prediksi Jumlah Peserta Mata Kuliah Mahasiswa Informatika Universitas Tanjungpura Menggunakan Regresi Linier Berganda," *Juwara: Jurnal Aplikasi dan Riset Informatika*, vol. 02, no. 1, pp. 14–25, 2023, doi: 10.26418/juara.v2i1.71853.

[9] N. R. Setyoningrum, P. J. Rahimma, S. T. Teknologi, I. Tanjungpinang, and K. Tanjungpinang, "Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang," in *SNISTEK 4*, 2022, pp. 13–19. Accessed: Mar. 30, 2024. [Online]. Available: https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/5200

[10] Z. Huang, Y. Mei, and M. Zhang, "Investigation of Linear Genetic Programming for Dynamic Job Shop Scheduling," in *2021 IEEE Symposium Series on*

Computational Intelligence, SSCI 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/SSCI50451.2021.9660091.

[11] M. Xu, Y. Mei, F. Zhang, and M. Zhang, "Genetic Programming with Lexicase Selection for Large-scale Dynamic Flexible Job Shop Scheduling," *IEEE Transactions on Evolutionary Computation*, p. 1, 2023, doi: 10.1109/TEVC.2023.3244607.

[12] M. H. Mehta, N. C. Chauhan, and A. Gokhale, "Predicting Institute Graduation Rate With Genetic Algorithm Assisted Regression For Education Data Mining Predicting Institute Graduation Rate With Genetic Algorithm Assisted Regression For Education Data Mining," *Ictact Journal on Soft Computing*, vol. 11, no. 2, pp. 2266–2278, 2021, doi: 10.21917/ijsc.2021.0324.

[13] F. de la Rosa-Rivera, J. I. Nunez-Varela, J. C. Ortiz-Bayliss, and H. Terashima-Marín, "Algorithm selection for solving educational timetabling problems," *Expert Syst Appl*, vol. 174, 2021, doi: 10.1016/j.eswa.2021.114694.

[14] D. N. Gujarati, *Basic Econometrics*, Fourth.

[15] C. Rozikin and A. Solichin, "Implementation of Genetic Algorithm and Multi-Linear Regression For Predicting Food Supplies At Fast Food Restaurants," in *Prosiding Seminar Nasional Multidisiplin Ilmu Universitas Budi Luhur*, Apr. 2017.

[16] S. C. Amdani and D. Permana, "Penerapan Algoritma Genetika Untuk Penyeleksian Variabel Pada Analisis Regresi Logistik Biner," *Jurnal Pendidikan Tambusai*, vol. 7, no. 2, pp. 3844–3853, 2023, Accessed: Mar. 30, 2024. [Online]. Available: https://jptam.org/index.php/jptam/article/view/6734

[17] F. Insani and I. Darlianti, "Pembentukan Model Regresi Linier Menggunakan Algoritma Genetika untuk Prediksi Parameter Indeks Standar Pencemar Udara (ISPU)," *Jurnal CoreIT*, vol. 5, no. 2, 2019, Accessed: Mar. 30, 2024. [Online]. Available: https://ejournal.uin-suska.ac.id/index.php/coreit/article/view/9157

[18] R. Zhang, S. Chen, Z. Zhang, and W. Zhu, "Genetic Algorithm in Multimedia Dynamic Prediction of Groundwater in Open-Pit Mine," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/8556103.