

Systematic Composite Macro Analysis and Prediction of Rural School Students Performance using Efficient Data Mining Techniques

Suresh Solomon.G¹, Nancy Jasmine Golden²,

¹Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu
Email: sureshsolomong@rediffmail.com

²Associate Professor, Department of Computer Applications, Sarah Tucker College, Tirunelveli, Tamil Nadu

Abstract

Education in rural areas is of great concern when an entire society of this country is to attain its fulfilment in all of its domain development goals. The slow growth of educational standards implementation affects the learning environment development and the students in these rural areas. In India there exists a lot of Rural areas in which the educational performance of the rural school students are inferior when compared it to the performance of the urban areas due to the lack of facilities, environment, income, employment opportunities, and exposure. This paper presents the systematic composite analysis and prediction of rural school student's performance using efficient data mining techniques. The educational performance data are further processed for improving the student education environment. The final approach of this paper focuses on the proper selection of data mining tool to handle analysis the rural educational data collections for the improvements in various aspects of education environment. In near future this paper will be extended for the development of a real time model to access the rural educational environment using data mining approaches.

Keywords—Data Mining, Rural, Classification, Education, Performance

I. INTRODUCTION

Educational Data Mining focuses on developing new tools and algorithms for discovering data patterns. EDM develops methods and applies techniques from statistics [10], machine learning, and data mining to analyse data collected during teaching and learning. EDM tests learning theories and informs educational practice [8]. Educational data mining is emerging as a research area with a suite of computational and psychological methods and research approaches for understanding how students learn. New computer-supported interactive learning methods and tools—intelligent tutoring systems, simulations[3], games—have opened up opportunities to collect and analyse student data[6], to discover patterns and trends in those data, and to make new discoveries and test hypotheses about how students learn.

Goals of EDM:

1. Predicting students' future learning behaviour by creating student models that incorporate such detailed information as students' knowledge[1], motivation, met cognition, and attitudes;
2. Discovering or improving domain models that characterize the content to be learned and optimal instructional sequences [2];
3. Studying the effects of different kinds of pedagogical support that can be provided by learning software; and
4. Advancing scientific knowledge about learning and learners through building computational models that incorporate models of the student, the domain, and the software's pedagogy [7].

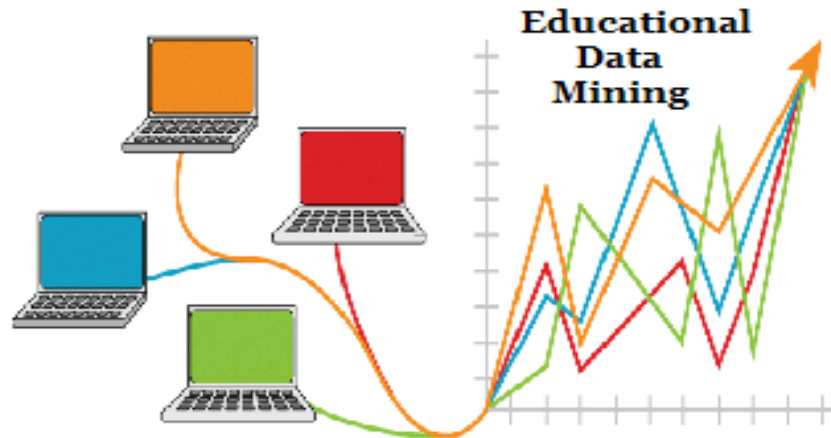


Fig-1: Educational Data Mining

II. LITERATURE REVIEW

- ❖ **Lei Tang, Xufei Wang, and Huan Liu –"Scalable Learning of Collective Behavior"** **IEEE 2016 Transaction on Knowledge and Data Engineering, Volume 34, Issue:6.**
This article deals with the statistical and numerical attributes and values available with the human society. It focuses on the demand and requirement satisfaction leads to the behavioral modifications in the society.
- ❖ **C. Romero, S. Ventura. "Educational Data Mining: A Review of the State-of-the-Art."** **IEEE Transaction on Systems, Man, and Cybernetics, 40(6), 601-618, 2016.**
This article reveals the concept of educational data mining with its features and impact in educational domain. The upcoming EDM model is the near focus for DM research entities.
- ❖ **M. Zaffar, M. A. Hashmani, and K. S. Savita, "Performance analysis of feature selection algorithm for educational data mining,"** **2017 IEEE Conf. Big Data Anal. ICBDA 2017, vol. 2018-Janua, pp. 7–12, 2018.**
This article discusses the concept of feature selection algorithms on educational data mining components with its performance analysis structure to improve the educational environment.
- ❖ **C. Jalota and R. Agarwal, "Analysis of Educational Data Mining using Classification,"** **Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Compute. Trends, Perspectives Prospect. Com. 2019, pp. 243–247, 2019.**
This article performs the analysis of educational data mining with its classification impact in educational domain using machine learning approaches; it also handles the categorization schema for educational domains for further processing.
- ❖ **P. Bachhal, S. Ahuja and S. Gargrish," Educational Data Mining: A Review",** **Journal of Physics: Conference Series, ICMAI-Mar-2021, doi:10.1088/1742-6596/1950/1/012022.**
This research article review data are extracted for analysis. Use of data derived from computational modeling for epidemiology and the application analysis for the various technical implementations. Address the complicated issue of data-driven science's limitations and connection to automated data retrievals.
- ❖ **P. Shabrina, B. Mostafavi, S. D. Tithi, M. Chi, and T. Barnes. Learning problem decomposition-recomposition with data-driven chunky parsons problem within an intelligent logic tutor. Pages 40–54, Bengaluru, India, July 2023. International Educational Data Mining Society.**
This article use a data driven graph-mining-based method to decompose historical student solutions of logic-proof problems into Chunks.

III.METHODOLOGY

The proposed methodology consists of 2 phases. The Problem - solution structure and architectural design.

a. The problem and solution structure:

The mathematical problem formulation for education process analysis is as follows,

i. **EPA-Education Process Analysis Score;** $0 \leq \text{EPA} \leq 1$ depends on the following fundamental components.

S-School facility

C-Curriculum (medium of instruction etc.)

T-Teacher qualification

TS-Teacher Student ratio

BG-Boys/Girls ratio

Compute $\text{EPA} = \text{Ep} = f(\text{Sa}, \text{Cb}, \text{Tc}, \text{TSd})$ such that

$f(\text{Sa}, \text{Cb}, \text{Tc}, \text{TSd}) = |\text{Sa}| + |\text{Cb}| + |\text{Tc}| + |\text{TSd}|$ where

$0 \leq |\text{Sa}| \leq 0.25,$

$0 \leq |\text{Cb}| \leq 0.25,$

$0 \leq |\text{Tc}| \leq 0.25,$

$0 \leq |\text{TSd}| \leq 0.25.$

The computation of Ep includes the following data mining approaches for each individual component.

S-School facility-Data mining Classification technique-C-curriculum-Data mining pattern analysis

T-Teacher qualification-Data mining Clustering techniques-Teacher Student ratio-Data mining

Decision trees approach.

ii.Student Performance Analysis:

The student performance analysis depends on the following basic 10 components.

For each individual student i ,

SPai-Student Physical ability/disability

SPmi-Student Motivation

SPfi-Student family economical background

SPri-Student family relation

SPti-Student school distance travel from home

SPvi-Student religion/caste variation effect

SPci-Student inter personnel communication

SPui-Student memory/understanding/language capability

SPoi-Student extracurricular interestis the primary objective

SPxi-Student exceptional behaviour/outputs

Compute $\text{SPi} = g(\text{SPai}, \text{SPmi}, \text{SPfi}, \text{SPri}, \text{SPti}, \text{SPvi}, \text{SPci}, \text{SPui}, \text{SPoi}, \text{SPxi})$

The computation of SPi comprises the fuzzy membership value allocation and computation using fuzzy logic approach.

iii.Student Performance Prediction:

The student performance prediction is based on the student performance analysis score with the following computation approach for each student i .

Student Performance Prediction score (SPPi)=Ep*SPi

Or

$SPPi=f(Sa,Cb,Tc,TSd) * g(SPai,SPmi,SPfi,SPri,SPti,SPvi,SPci,SPui,SPoi,SPxi)$ with the following interpretations as in table-1.

Table-1: Students Performance Prediction interpretation

Sl.No	SPPi	Prediction interpretation
1	>0.5	Improvement
2	=0.5	Flat average
3	<0.5	Non-improvement

The student performance prediction will be completed through data mining linear regression based approach.

b. Proposed methodology architecture design:

The proposed methodology architectural design is as follows in figure-2.

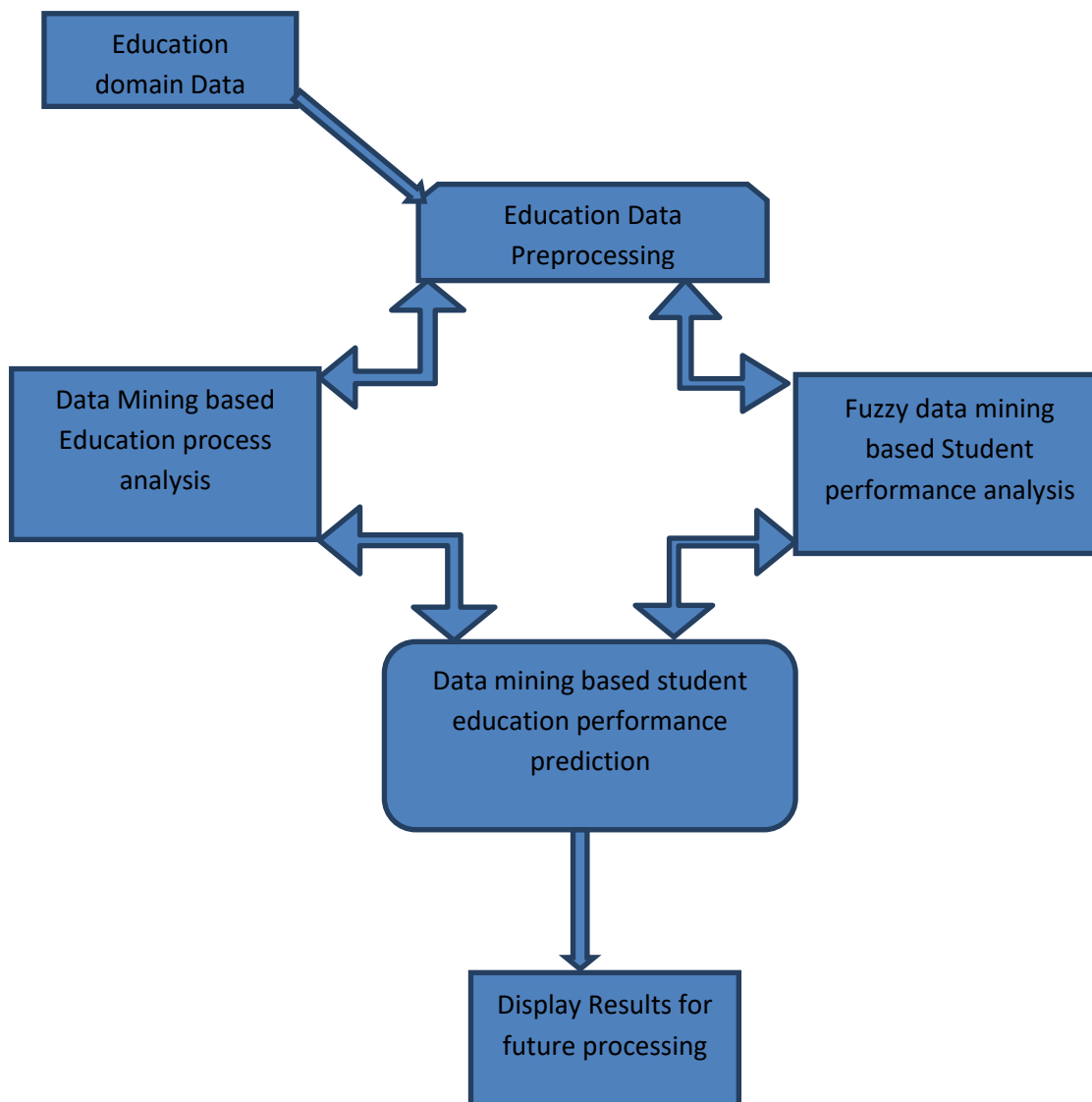


Fig-2:Proposed methodology architectural design for education performance prediction

IV. IMPLEMENTATION

The implementation consists of 5 stages, first 4 stages for education process analysis and 5th stage for student performance analysis using fuzzy logic approach.

a. Data mining Classification technique based School facility analysis-S

The school facility includes the features available for the students to enhance the education performance in an optimized ways in table-2.

The sample java code for dataclassification is as follows,

```
class Schoolfacclassify implements ValueBasedClassifier {
    public boolean matchValue(Object value) {
        if (value instanceof Number) {
            int intValue = ((Number)value).intValue();
            return intValue;
        }
    }
}
```

Table-2: School facility classification values

Sl.No	School facility	Classification	Classification value
1	Building, Library, Lab, Smart classroom, Extracurricular activity	A	0.25
2	Building, Library, Lab, Smart classroom	B	0.2
3	Building ,Library, Lab	C	0.15
4	Building, Library	D	0.1
5	Building	E	0.05

b. Data mining pattern analysis based curriculum structure value-C

The data mining pattern analysis based curriculum structure defines the languages such as Mother tongue, English, National/Foreign language, skills such as dance,music,martial arts,swimming etc. along with games,weekly technical events and cultural as in table-3.

Table-3: Students curriculum structure value-C

Sl.No	Curriculum Structure	Pattern	Curriculum pattern value
1	Languages-3, Skills, Games, Weekly events, Projects	Well defined	0.25
2	Languages-2. skills, games, cultural	Standard	0.2
3	Languages-2, games, cultural	Normal	0.15
4	Languages-2, games	Average	0.1
5	Languages-1 or 2	Below average	0.05

c. Data mining Clustering technique based Teacher qualification analysis-T

The teacher qualification includes the principal,UG,PG, P.Ed; Skill Teachers along with the medical and counselling assistants, the corresponding cluster with its values are represented in table-4.

Table-4: Teacher qualification analysis-T

Sl.No	Teacher Qualification	Cluster Label	Cluster value
1	Qualified Principal	A	0.05
2	Sufficient UG Teachers	B	0.05
3	Sufficient PG Teachers	C	0.05
4	Sufficient Language Teachers	D	0.05
5	Librarian	E	0.05
6	Physical education teachers	F	0.05
7	Arts/Skill development Trainer/Teacher	G	0.05
8	Guidance and Counselling Faculty	H	0.05
9	Medical Assistant faculty	I	0.05
10	Assistant teacher/technicians/non-teaching faculties	J	0.05
Total cluster value			0.25

d. Data mining Decision trees approach based Teacher Student ratio analysis-TS

The teacher student ratio values are assigned with the implementation of decision trees based approach as follows,

If Teacher: student ratio <= 1:10 then

TS=0.25; break;

Else if Teacher: student ratio <= 1:20 then

TS=0.2; break;

Else if Teacher: student ratio <= 1:30 then

TS=0.15; break;

Else if Teacher: student ratio <= 1:40 then

TS=0.1; break;

Else

TS=0.05; break;

e. Fuzzy computation for Student performance analysis-SPi

The fuzzy membership value allocation for the student performance analysis is represented as follows in the table 5.

Table-5: Student performance analysis-SPi

Sl.No	Factor Name	Fuzzy membership value for positive response
1	SPai-Student Physical ability/disability	0.1
2	SPmi-Student Motivation	0.1
3	SPfi-Student family economical background	0.1
4	SPri-Student family relation	0.1
5	SPti-Student school distance travel from home	0.1
6	SPvi-Student religion/caste variation effect	0.1
7	SPci-Student inter personnel communication	0.1
8	SPui-Student memory/understanding/language capability	0.1
9	SPoi-Student extracurricular interest is the primary objective	0.1
10	SPxi-Student exceptional behaviour/outputs	0.1
Total		1.0

V. RESULTS AND DISCUSSION

The results of 12 educational institution in and around Tenkasi and Tuticorin district, Tamilnadu, India are collected through frequent observations as mentioned in table-6 is as follows,

Table-6: Educational institution data collection

School /student data of Educational institution type	Count
Private	3
Aided	4
Government	5

The sample data collection sheets are represented in the following figure-3,

Name of the School: Govt. Hr. Sec. School
 Govt./Aided/Private: Kambankulam.

The School

No. of Buildings	2
School timings	9.10 to 4.20 Pm
School Principal/Headmaster	B. RUBAN SELVA KUMAR
School Administrators	HM

The curriculum (intended and Implemented)

Language of instruction	Tamil
Other Languages	ENGLISH

Teachers

Level of education	UG BEd - 7
UG B.Ed. and PG B.Ed. count	PG BEd - 9
Pre-service training count	Not Applicable.
In-service training count	4
Lesson Plan Maintained status	Well Maintained.

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	17	16	17	16	16	20	29
Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	6	4	5	11	6	19	13

Total student count: 128
 Failures in each class count: 100% Pass.

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	NIL	NIL	NIL	NIL	NIL	NIL	NIL

Number per class (teacher-student ratio)

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	1:40	1:40	1:40	1:40	1:40	1:40	1:40

Teaching-Learning Materials: 150

Library Books count: 800 Books
 Projector count: 01
 Exhibition visit count: 04

[Signature] 25/02/2021
 B. RUBAN SELVA KUMAR
 Govt. Hr. Sec. School
 Kambankulam
 Tamil Nadu - 628704

Name of the School: Govt. Hr. Sec. School
 Govt./Aided/Private: Kambankulam.

The School

No. of Buildings	2
School timings	9.10 to 4.20 Pm
School Principal/Headmaster	B. RUBAN SELVA KUMAR
School Administrators	HM

The curriculum (intended and Implemented)

Language of instruction	Tamil
Other Languages	ENGLISH

Teachers

Level of education	UG BEd - 7
UG B.Ed. and PG B.Ed. count	PG BEd - 9
Pre-service training count	Not Applicable.
In-service training count	4
Lesson Plan Maintained status	Well Maintained.

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	17	16	17	16	16	20	29
Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	6	4	5	11	6	19	13

Total student count: 128
 Failures in each class count: 100% Pass.

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	NIL	NIL	NIL	NIL	NIL	NIL	NIL

Number per class (teacher-student ratio)

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	1:40	1:40	1:40	1:40	1:40	1:40	1:40

Teaching-Learning Materials: 150

Library Books count: 800 Books
 Projector count: 01
 Exhibition visit count: 04

[Signature] 25/02/2021
 B. RUBAN SELVA KUMAR
 Govt. Hr. Sec. School
 Kambankulam
 Tamil Nadu - 628704

Name of the School: Govt. High School
 Govt./Aided/Private: PUTHUKULAM.

The School

No. of Buildings	1
School timings	9.10 am to 4.20 P.M
School Principal/Headmaster	Headmaster
School Administrators	Headmaster, S. Mariya Jaha Britto

The curriculum (intended and Implemented)

Language of instruction	Tamil
Other Languages	NIL

Teachers

Level of education	08
UG B.Ed. and PG B.Ed. count	Not applicable
Pre-service training count	03
In-service training count	03
Lesson Plan Maintained status	Well Maintained.

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	5	5	9	7	11	-	-
Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	7	5	2	9	8	-	-

Total student count: 68
 Failures in each class count: 100% pass

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	NIL	NIL	NIL	NIL	NIL	-	-

Number per class (teacher-student ratio)

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	1:40	1:40	1:40	1:40	-	-	-

Teaching-Learning Materials

Library Books count: 500
 Projector count: only one
 Exhibition visit count: 01

[Signature]
 Headmaster
 Govt. High School
 Puthukulam - 628704
 Thoothukudi Dist

Name of the School: Government High Secondary School
 Govt./Aided/Private: Maruthampuram, Alagunathan Taluk - 627851

The School

No. of Buildings	13
School timings	9.30 AM - 1.30 P.M
School Principal/Headmaster	01
School Administrators	Govt.

The curriculum (intended and Implemented)

Language of instruction	02 (Tamil, English)
Other Languages	-

Teachers

Level of education	PG - 8, BT - 17, SGT - 4, P.E.T - 1, Dooking - 1
UG B.Ed. and PG B.Ed. count	Total - 32
Pre-service training count	25
In-service training count	25
Lesson Plan Maintained status	YES

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	55	53	63	107	86	75	60
Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	43	60	43	111	101	60	100

Total student count: 1017
 Failures in each class count: As per RTE-2009 norms 3-6 students = 01 Teachers

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	3	3	3	6	5	9	9

Number per class (teacher-student ratio): As per RTE-2009 norms 3-6 students = 01 Teachers

Teaching-Learning Materials: 61-90 (for 20 students) = 01 Teachers

Library Books count: 500
 Projector count: 01
 Exhibition visit count: 02

[Signature]
 HEADMASTER
 Govt. Hr. Sec. School
 MARUTHAMPURAM
 D.O Code - SB 408

Name of the School: **Government Higher Secondary School, Alangulam**
 Govt./Aided/Private: **Govt.**

The School

No. of Buildings	12
School timings	9.30 AM - 4.30 PM
School Principal/Headmaster	1
School Administrators	Govt.

The curriculum(Intended and Implemented)

Language of instruction	2 (Tamil, English)
Other Languages	-

Teachers

Level of education	PG-13 BT-20, S.G.T-9, P.E.T.-2
UG B.Ed. and PG B.Ed. count	Dravuni created Total-46
Pre-service training count	33
In-service training count	33
Lesson Plan Maintained status	Yes

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	53	54	59	95	126	123	124

Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	63	64	66	90	114	105	136

Total student count: 1292

Failures in each class count													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	-	-	-	-	-	-	-

Number per class(teacher-student ratio) As per RTE-2009 Norms 1-60 students = 01 Teacher

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	3	3	4	6	7	7	4

Teaching-Learning Materials

Library Books count: 1000
 Projector count: 02
 Exhibition visit count: 01

J. Frank W. Watson
 01/02/2021
 HEAD MISTRESS
 GOVT. HIGHER SECONDARY SCHOOL,
 ALANGULAM - 627851
 THOOTHUKUDI DIST.

Name of the School: **Solomon Matric. Hr. Sec. School, Nazareth**
 Govt./Aided/Private: **Private**

The School

No. of Buildings	3
School timings	8.30 - 4.20
School Principal/Headmaster	1
School Administrators	1

The curriculum(Intended and Implemented)

Language of instruction	2
Other Languages	-

Teachers

Level of education	21
UG B.Ed. and PG B.Ed. count	21
Pre-service training count	21
In-service training count	21
Lesson Plan Maintained status	Yes

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
6	20	20	27	21	39	28	30	30	20	11	22	20	14

Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
9	31	33	26	14	27	31	24	24	25	22	24	19	10

Total student count: 637

Failures in each class count													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	-	-	-	-	-	-	-

Number per class(teacher-student ratio)

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
2	2	2	2	2	2	2	2	2	2	2	2	3	2

Teaching-Learning Materials

Library Books count: 1500
 Projector count: -
 Exhibition visit count: -

J. Frank W. Watson
 18/02/2021
 PRINCIPAL
 SOLOMON MATRICULATION SCHOOL
 JAMIN NAGAR
 NAZARETH - 627817

Name of the School: **JAMES MEMORIAL MATRIC. HR. SEC. SCHOOL, PRAKASAPURAM, MUKUPERI - (PO), PIN CODE- 62864**
 Govt./Aided/Private: **Private**

The School

No. of Buildings	5
School timings	9.20 a.m - 4 Pm
School Principal/Headmaster	01
School Administrators	PRIVATG

The curriculum(Intended and Implemented)

Language of instruction	02 (Tamil, English)
Other Languages	-

Teachers

Level of education	PG-8, BT-10, Lang-4
UG B.Ed. and PG B.Ed. count	P.E.T-1
Pre-service training count	-
In-service training count	-
Lesson Plan Maintained status	YES

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
5	18	9	10	9	16	9	14	13	14	16	15	28	22

Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
6	14	9	10	5	10	10	8	7	7	4	7	24	20

Total student count: 337

Failures in each class count													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	-	-	-	-	-	-	-

Number per class(teacher-student ratio)

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	2	1	1	2	1	1	1

Teaching-Learning Materials

Library Books count: 2500
 Projector count: 01
 Exhibition visit count: 01

J. Frank W. Watson
 01/02/2021
 PRINCIPAL
 JAMES MEMORIAL MATRICULATION SCHOOL
 HIGHER SECONDARY SCHOOL OF B.D.A.
 PRAKASAPURAM - 62864
 THOOTHUKUDI DIST. (TN)

Name of the School: **St Joseph's matric higher secondary school Alangulam - 627851**
 Govt./Aided/Private: **Private**

The School

No. of Buildings	3
School timings	9.00 am to 4.30 P.M
School Principal/Headmaster	01
School Administrators	matriculation

The curriculum(Intended and Implemented)

Language of instruction	01 (English)
Other Languages	-

Teachers

Level of education	PG-13, BT-10, S.G.T-8, P.E.T-1
UG B.Ed. and PG B.Ed. count	Total : 15
Pre-service training count	-
In-service training count	-
Lesson Plan Maintained status	YES

Students

Boys count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
39	48	73	78	75	47	61	44	46	41	42	43	28	31

Girls count/class													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
59	53	59	44	61	51	65	32	40	37	41	37	40	33

Total student count: 1348

Failures in each class count													
LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
-	-	-	-	-	-	-	-	-	-	-	-	-	-

Number per class(teacher-student ratio)

LKG	UGK	1	2	3	4	5	6	7	8	9	10	11	12
3	3	4	3	4	3	3	2	2	2	2	2	2	3

Teaching-Learning Materials

Library Books count: 500
 Projector count: 1
 Exhibition visit count: -

Santhosh
 18/02/2021
 PRINCIPAL
 St. Joseph's Mat. Hr. Sec. School,
 Adakkala Nagar, Alangulam-627851.

Fig-3: Real-time data collection sheets from Schools

The proposed methodology implementation for the analysis and prediction of students performance results by taking each student sample from all the 12 schools are as follows in Table-7,

Table-7: Student performance prediction score computation

No	Student ID	School facility analysis-S	curriculum structure value-C	Teacher qualification analysis-T	Teacher Student ratio analysis-TS	Education process $E_p=f(S_a,C_b,T_c,T_s)$	Student performance analysis-SPI	Student Performance Prediction score (SPPi)= $E_p * S_{PI}$
1	P1S1	0.25	0.25	0.25	0.25	1	0.8	0.8
2	P2S2	0.25	0.25	0.25	0.2	0.95	0.8	0.76
3	P3S3	0.2	0.2	0.2	0.2	0.8	0.7	0.56
4	A1S1	0.25	0.25	0.25	0.2	0.95	0.8	0.76
5	A2S2	0.25	0.25	0.2	0.2	0.9	0.8	0.72
6	A3S3	0.2	0.2	0.2	0.2	0.8	0.8	0.64
7	A4S4	0.2	0.2	0.2	0.2	0.8	0.7	0.56
8	G1S1	0.25	0.25	0.25	0.2	0.95	0.8	0.76
9	G2S2	0.25	0.25	0.2	0.2	0.9	0.7	0.63
10	G3S3	0.25	0.25	0.2	0.2	0.9	0.6	0.54
11	G4S4	0.2	0.2	0.15	0.15	0.7	0.6	0.42
12	G5S5	0.2	0.15	0.15	0.15	0.65	0.5	0.325

The performance analysis and prediction of student education performance shows that 10 out of 12 students will perform well in their current education state.

While comparing the proposed methodology results with the quarterly and half yearly performance analysis the final results shows that the proposed methodology prediction results are accurate with 11 out of 12 student's performance.

The proposed methodology produces 91.67% of success when implemented directly with the real time school environment in and around Tenkasi and tucorin districts.

The following graph shows the final results for the proposed methodology implementation in real-time environment.

Results for Systematic Composite Analysis and Prediction of Rural School Students Performance using Efficient Data Mining Techniques

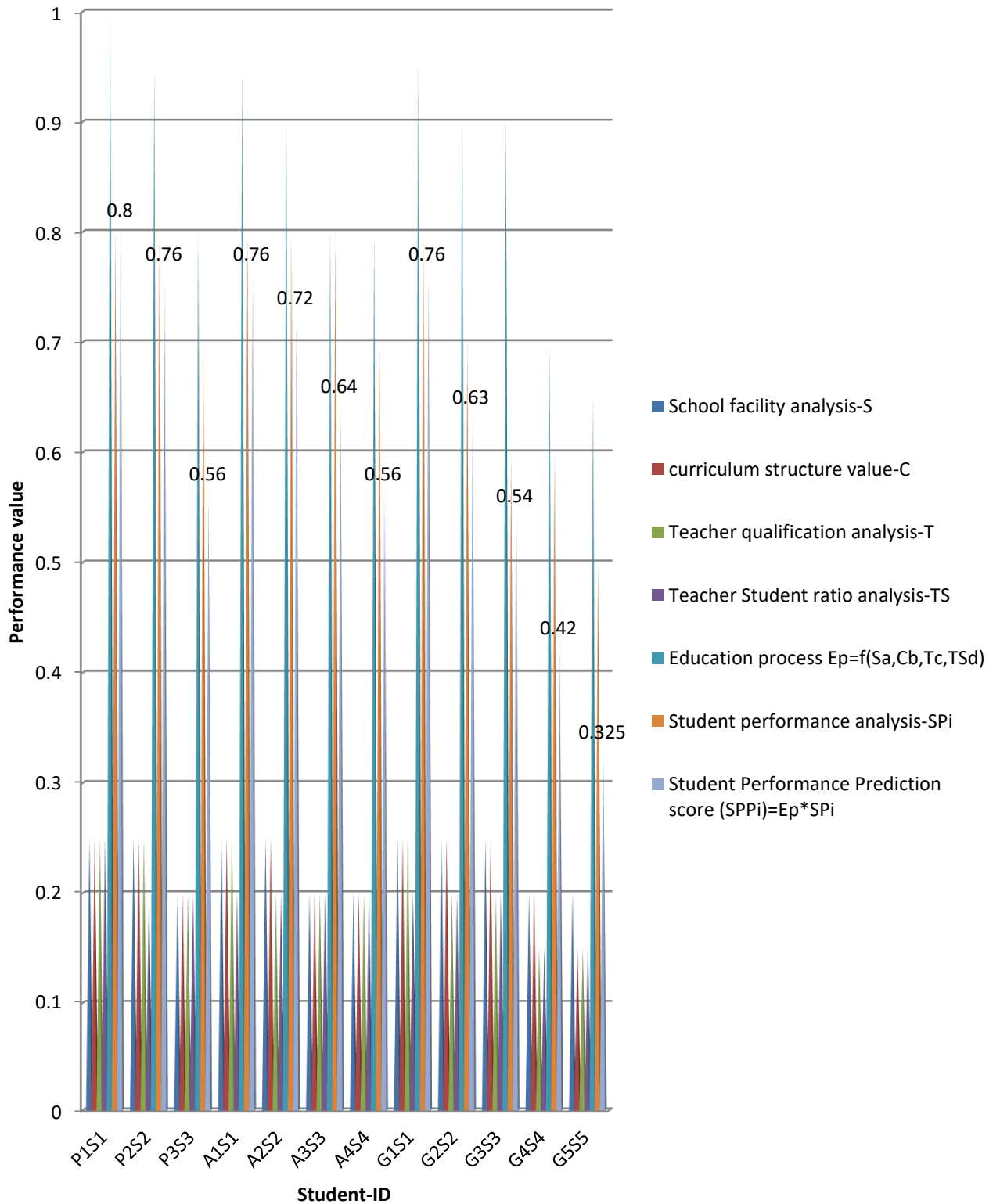


Fig-4: Student performance prediction score computation result

The comparison of the proposed methodology with the existing approaches along with the computation of accuracy, recall, precision and F1 score is as follows in table-8,

Table-8: Proposed methodology comparison using accuracy, precision, recall and F1 score

Sl.No	Methodology	Accuracy	Precision= $\frac{tp}{(tp+fp)}$	Recall= $\frac{tp}{(tp+fn)}$	F1 score= $\frac{2*tp}{(2*tp+fp+fn)}$
1	Node detection method	69%	0.68	0.72	0.77
2	Trend extrapolation	72%	0.71	0.76	0.81
3	Back casting	78%	0.76	0.8	0.84
4	Proposed systematic composite macro analysis and prediction of rural school students performance using efficient data mining techniques	92%	0.91	0.93	0.94

VI. CONCLUSION

Education data mining plays an important role in the field of education environment. The process of handling educational data for the welfare of student education improvement system is an essential criterion for the development of any country. This research article emphasizes on systematic composite analysis and prediction of rural school students performance using efficient data mining techniques. The initial phase of this research module focuses on the student data collection and school facilities information collection. Then this research article performs the education process analysis score computation using the educational data mining techniques such as classification, pattern analysis, and clustering and decision trees. The final module focuses on the student performance analysis computation using the fuzzy logic based membership value assignment. The final student performance prediction score computation depends on the education process score and student performance analysis score. The practical implementation of this proposed research methodology produces exact results for 11 out of 12 students education performance prediction score. In future we will implement neural networks and genetic approach based methodology for faster and accurate results in this educational data mining domain.

REFERENCES

- [1] M. Newman, —Finding community structure in networks using the eigenvectors of matrices, *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 74, no. 3, 2006. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevE.74.036104>
- [2] Drachsler, H., Hummel, H. G., Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. In *International Journal of Learning Technology*. 3,4, 404-423.
- [3] Khribi, M. K., Jemni, M., Nasraoui, O. (2008). Automatic Recommendations for ELearning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In *IEEE International Conference on Advanced Learning Technologies*, Washington, DC, 241-245.
- [4] L. Tang and H. Liu, —Toward predicting collective behaviour via social dimension extraction, *IEEE Intelligent Systems*, vol. 25, pp. 19–25, 2013.

- [5] Baker, Ryan. *Data Mining for education*, oxford, UK: Elsevier. Retrieved 9, February 2014.
- [6] Lei Tang, Xufei Wang, and Huan Liu –"Scalable Learning of Collective Behavior" IEEE 2016 Transaction on Knowledge and Data Engineering, Volume 34.Issue:6.
- [7] C. Romero, S. Ventura. Educational Data Mining: A Review of the State-of-the-Art. IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 40(6), 601-618, 2016.
- [8] M. Zaffar, M. A. Hashmani, and K. S. Savita, —Performance analysis of feature selection algorithm for educational data mining, 2017 IEEE Conf. Big Data Anal. ICBDA 2017, vol. 2018-Janua, pp. 7–12, 2018.
- [9] C. Jalota and R. Agrawal, —Analysis of Educational Data Mining using Classification, Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019, pp. 243–247, 2019.
- [10] P. Bachhal, S. Ahuja and S. Gargrish, "Educational Data Mining: A Review", Journal of Physics: Conference Series, ICMAI-Mar-2021, doi:10.1088/1742-6596/1950/1/012022
- [11] P. Shabrina, B. Mostafavi, S. D. Tithi, M. Chi, and T. Barnes. Learning problem decomposition-recomposition with data-driven chunky parsons problem within an intelligent logic tutor. In M. Feng, T. Kaser, and P. Talukdar, editors, Proceedings of the 16th International Conference on Educational Data Mining, pages 40–54, Bengaluru, India, July 2023.