A

Project file on

ANALYSIS OF STUDENT BEHAVIOUR UNDER NATIONAL EDUCATION POLICY 2020 USING MACHINE LEARNING TECHNIQUE

Bachelor of Technology

Computer Science & Information Technology

By

Ashutosh Yadav(21CS34)

Under Supervision of

Dr. Iram Naim

चरैवेति चरैवेति

महात्मा ज्योतिबा फुले
रुहेलखण्ड विश्वविद्यालय, बरेली

Department of Computer Science & Information Technology

Faculty of Engineering & Technology

M.J.P. Rohilkhand University, Bareilly

2024 – 2025

ABSTRACT

The National Education Policy (NEP) 2020 has introduced transformative changes to India’s education system, aiming to enhance student learning outcomes and foster holistic development. This project, titled **"Analysis of Student Behaviour Under National Education Policy 2020 Using Machine Learning Techniques**" seeks to understand the behavioral patterns of students within the framework of NEP 2020 and evaluate its impact on their academic and personal growth.

Using machine learning models, the project analyzes data collected from various sources, such as surveys, academic records, participation in extracurricular activities, and feedback from educators and students. Key objectives include identifying trends in student engagement, adaptability to skill-based education, and the efficacy of policy-driven curriculum changes.

The study employs techniques like clustering, classification, and sentiment analysis to uncover insights into how students respond to new assessment patterns, multidisciplinary approaches, and experiential learning initiatives. Predictive analytics is also utilized to forecast long-term outcomes, such as career readiness and the development of critical thinking skills.

By providing data-driven insights, this research aims to guide educators and policymakers in fine-tuning NEP 2020 implementation strategies, ensuring that the policy's objectives are met effectively. The findings will contribute to shaping a more adaptive, inclusive, and outcome-oriented education system.

ACKNOWLEDGEMENT

We are extremely grateful and remain indebted to our mentor , **Dr. Iram Naim**, for being a source of inspiration and for their constant support in the design, Implementation and evaluation of the project We are thankful to them for their constant constructive criticism and valuable suggestions, which lead us to a beneficial stage while developing the project on "**Analysis of student behaviour under national education policy 2020 using machine learning technique**". They have been a constant source of inspiration and motivation for hard work. They have been very cooperative. throughout this project's work. Through this column, it would be our utmost pleasure to express our warm thanks to them for their encouragement, cooperation, and consent without which we might not be able to accomplish this project. We also express our gratitude to **Dr. Iram, Naim** for providing us the infrastructure to carry out the project and to all staff members who were directly and indirectly instrumental in enabling us to stay committed to the project. Lastly, we would like to thank our team members, without their valuable suggestions, we cannot complete our project.

Ashutosh Yadav(220089020036)

INDEX

1.INTRODUCTION

India’s National Education Policy (NEP) 2020 envisions higher education as a key driver of individual and societal well-being, fostering creativity, ethics, social responsibility, and 21st-century skills. It aims to address challenges like fragmented education, rigid specialization, limited access, lack of autonomy, and insufficient research funding. Key reforms include establishing multidisciplinary universities, promoting local language instruction, curriculum and assessment improvements, and creating a National Research Foundation. Governance reforms include merit-based appointments and a streamlined regulatory system. The goal is to create an inclusive, innovative education system that supports individual growth and national development.

The National Education Policy (NEP) 2020 emphasizes the transformative role of higher education in India’s societal and economic development. It aims to foster well-rounded, ethical, and creative individuals with 21st-century skills. Key reforms include transitioning to large, multidisciplinary institutions, promoting local language instruction, enhancing faculty autonomy, and improving governance. The policy seeks to address challenges like limited access, fragmented systems, lack of focus on cognitive skills, and inadequate research funding. By prioritizing equity, inclusion, and holistic education, NEP 2020 envisions a more vibrant, socially engaged, and innovative India.

To Identify these changes in the higher education by new education policy there is a need to analyse suitable combination of subjects during there course like undergraduate . To analyse these combination of subjects we will implement machine learning based model(Association rule mining) to our university dataset . Our university Mahatma Jyotiba Phule Rohilkhand University is adopted new education policy in the year 2022 . More than 463816 undergraduate students are currently studying in the university. In this project we will study about various student related details and based upon that try to find out suitable combination of subjects from the own department and other department of university .

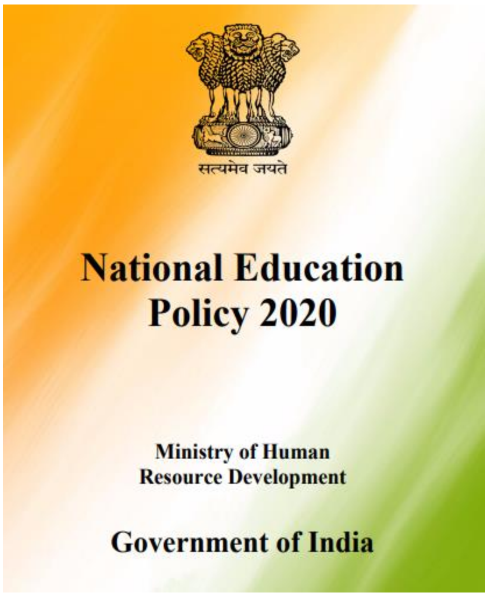
1.1 CONTEXT

Association rule learning is a fundamental data mining technique used to discover interesting relationships, patterns, or associations among a set of items in large datasets. It is widely applied in various fields, including market basket analysis, recommendation systems, and customer behavior analysis.

1.2 MOTIVATION

The motivation behind association rule learning lies in uncovering hidden patterns and relationships within large datasets. By identifying item associations, businesses can optimize decision-making, enhance customer experiences, and improve marketing strategies, such as product recommendations or promotions, based on insights drawn from customer behavior and transactional data.

2.NATIONAL EDUCATION POLICY 2020



The National Education Policy (NEP) 2020 is a landmark framework introduced by the Government of India to transform the education system, making it more holistic, flexible, and aligned with the needs of the 21st century. It envisions an education system rooted in Indian values while being globally competitive. The policy introduces significant changes in school and higher education, focusing on universal access, equity, quality, affordability, and accountability.

In school education, NEP replaces the traditional 10+2 structure with a 5+3+3+4 design, tailored to the developmental stages of children. Early childhood education is emphasized, and teaching in regional languages or mother tongues is encouraged, particularly in the foundational years. The curriculum is designed to focus on core concepts, critical thinking, and experiential learning, moving away from rote memorization. Board examinations are reformed to assess understanding and application rather than memory.

In higher education, NEP aims to increase enrollment and provide greater flexibility with a multidisciplinary approach. It allows students to enter and exit programs at different stages, earning certifications or degrees accordingly. The policy emphasizes research and innovation, promoting the establishment of a National Research Foundation. Vocational education is integrated into mainstream education, starting from an early stage, to develop practical skills and employability.

Teachers' roles are redefined, with emphasis on rigorous training and continuous professional development. Technology is integrated into teaching and learning, enabling access to high-quality digital resources and fostering skills like coding and artificial intelligence. Overall, NEP 2020 seeks to create an education system that is inclusive, dynamic, and adaptable to the evolving needs of society, with a long-term goal of full implementation by 2040.

Preparation of professionals must involve an education in the ethic and importance of public purpose, an education in the discipline, and an education for practice. It must centrally involve critical and interdisciplinary thinking, discussion, debate, research, and innovation. For this to be achieved, professional education should not take place in the isolation of one's specialty. 20.2. Professional education thus becomes an integral part of the overall higher education system. Stand-alone agricultural universities, legal universities, health science universities, technical universities, and stand-alone institutions in other fields, shall aim to become multidisciplinary institutions offering holistic and multidisciplinary education. All institutions offering either professional or general education will aim to organically evolve into institutions/clusters offering both seamlessly, and in an integrated manner by 2030. 20.3. Agricultural education with allied disciplines will be revived. Although Agricultural Universities comprise approximately 9% of all universities in the country, enrolment in agriculture and allied sciences is less than 1% of all enrolment in higher education. Both capacity and quality of agriculture and allied disciplines must be improved in order to increase agricultural productivity through better skilled graduates and technicians, innovative research, and market-based extension linked to technologies and practices. The preparation of professionals in agriculture and veterinary sciences through programmes integrated with general education will be increased sharply. The design of agricultural education will shift towards developing professionals with the ability to understand and use local knowledge, traditional knowledge, and emerging technologies while being cognizant of critical issues such as declining land productivity, climate change, food sufficiency for our growing population, etc. Institutions offering agricultural education must benefit the local community directly; one approach could be to set up Agricultural Technology Parks to promote technology incubation and dissemination and promote sustainable methodologies. 20.4. Legal education needs to be competitive globally, adopting best practices and embracing new technologies for wider access to and timely delivery of justice. At the same time, it must be informed and illuminated with Constitutional values of Justice - Social, Economic, and Political - and directed towards national reconstruction through instrumentation of democracy, rule of law, and human rights. The curricula for legal studies must reflect socio-cultural contexts along with, in an evidence-based manner, the history of legal thinking, principles of justice, the practice of jurisprudence, and other related content appropriately and adequately. State institutions offering law education must consider offering bilingual education for future lawyers and judges - in English and in the language of the State in which the institution is situated. 20.5. Healthcare education needs to be re-envisioned so that the duration, structure, and design of the educational programmes need to match the role requirements that graduates will play. Students will be assessed at regular intervals on well-defined parameters primarily required for working in primary care and in secondary hospitals. Given that people exercise pluralistic choices in healthcare, our healthcare education system must be integrative meaning thereby that all students of allopathic medical education must have a basic understanding of Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homeopathy (AYUSH), and vice versa. There shall also be a much greater emphasis on preventive healthcare and community medicine in all forms of healthcare education. National Education Policy 2020 51 20.6. Technical education includes degree and diploma programmes in, engineering, technology, management, architecture, town planning, pharmacy, hotel management, catering technology etc., which are critical to India 's overall development. There will not only be a greater demand for wellqualified manpower in these sectors, it will also require closer collaborations between industry and higher education institutions to drive innovation and research in these fields. Furthermore, influence of technology on human endeavours is expected to erode the silos between technical education and other disciplines too. Technical education will, thus, also aim to be offered within multidisciplinary education institutions and programmes and have a renewed focus on opportunities to engage deeply with other disciplines. India must also take the lead in preparing professionals in cutting-edge areas that are fast gaining prominence, such as Artificial Intelligence (AI), 3-D machining, big data analysis, and machine learning, in addition to genomic studies, biotechnology, nanotechnology, neuroscience, with important applications to health, environment, and sustainable living that will be woven into undergraduate education for enhancing the employability of the youth.

3.OBJECTIVE

The primary objective of association rule learning is to discover meaningful patterns and relationships between items in large datasets. It aims to identify frequently occurring item sets and generate rules that can predict the co-occurrence of items.

The main objective of project to find the association relation between subjects paper code of B.A course by analysing data of student course subject. The apriori algorithm helps uncover the frequent subject combination, offering insights into popular course path ways. This can aid academic planners and students in making inform decision about course offerings, creating more efficient academic schedules, and providing personalized recommendation for future course selections based on observed student preferences.

The objective is to evaluate the impact of NEP 2020 on student engagement, learning outcomes, and holistic development. The project leverages machine learning techniques like association rule technique to analyze data from academic records, surveys, and feedback. These techniques help uncover patterns in student participation, adaptability, and critical thinking skills. Predictive modeling is also used to forecast long-term outcomes, such as career readiness and the ability to acquire new skills

3.1Brief Description

This project aims to analyze student behavior and learning outcomes under the framework of the National Education Policy (NEP) 2020, using advanced machine learning (ML) techniques. The NEP 2020 introduced a transformative vision to revamp India's education system by emphasizing holistic, multidisciplinary, and skill-oriented learning.

The apriori algorithm can be applied to mine association rule from data on student course enrollments. The goal is to identify frequent pattern in the course that’s students tend to take together. For example, the algorithm may reveal that student who enroll in “History” obtain also choose “polity” or “statics” , or “economic”. By discovering these association, the algorithm provides valuable insights that can help academic planners optimize course offering structure timetable, and make personalized recommendations for students based on their past behaviors. These can enhance student satisfaction and improve the overall efficiency of course scheduling.

4.DATA DESCRIPTION

In this project the given dataset contain all of the data over whole university students , this dataset contain many of the important attributes . This dataset is real time world dataset , including all of the information of student with their name , courses and results and category and other important informations . This given dataset for the“**Analysis of Student Behaviour Under National Education Policy 2020 Using Machine Learning Techniques**” competition is part of the annual Machine learning and Data Science survey . This dataset is unique because it provides raw, anonymized data, allowing for detailed exploration without privacy concerns.

Dataset Overview

The dataset provided contains detailed academic information about students enrolled in a specific program. It is structured to represent student records, focusing on their enrollment details and performance across several subjects or courses. Each row represents a unique student, while the columns capture various attributes related to their academic journey.

Structure of the Dataset

The dataset is stored in a single sheet titled **"student"** and comprises 11 columns. These columns provide information about the students' class identification, enrollment details, names, and the numeric identifiers for the subjects or courses they are taking in the semester. Below is a breakdown of the columns and their likely significance:

- **classid**
  This column indicates the class and semester of the student. For instance, "BA3SEM6" suggests that the student is in the third year of the Bachelor of Arts program, attending the sixth semester. This field helps categorize students by their academic progression.
- **enrol**
  This column represents the enrollment number, a unique identifier assigned to each student. It ensures individual records can be tracked without ambiguity and likely serves as the primary key in this dataset.
- **name**
  The "name" column contains the full name of the student. It provides a

personal identifier, allowing the dataset to be more human-readable and meaningful.

- **Paper1 to Paper8**

These columns represent numeric codes for the subjects or courses taken by each student during the semester. The exact nature of these codes is not provided but could correspond to course IDs in the university's academic catalog. The codes appear to follow a consistent format, which may indicate systematic categorization based on the program structure.

Sample Data

Here is a snapshot of the dataset to help visualize its structure:

classid	enrol	name	paper1	paper2	paper3	paper4	paper5	paper6	paper7	paper8
BA3SEM6	21053226	AARTI DEVI	31423	32907	41427	41428	42907	42908	49006	49032
BA3SEM6	21053228	AFSHEEN BEE	31602	32805	41607	41608	42807	42808	49006	49032
BA3SEM6	21053232	AKANKSHA TRIVEDI	32702	32805	42709	42710	42807	42808	49006	49032
BA3SEM6	21053234	AMISHA GANGWAR	32402	32702	42408	42409	42709	42711	49006	49032
BA3SEM6	21053236	ANJALI	32402	32805	42408	42409	42807	42808	49006	49032

Detailed Column Descriptions

1. classid

The "classid" field provides insight into the academic structure of the program. For example:

- "BA" suggests the Bachelor of Arts program.

- The number "3" likely indicates the year of study (e.g., 3rd year).
- "SEM6" indicates the semester (e.g., 6th semester).

This classification is useful for filtering students by their academic level and grouping them by year or semester. It can also help in identifying curriculum patterns for a specific cohort.

2. enrol

The enrollment number serves as a unique identifier for each student. This field is vital for:

- Linking individual performance data to other records (e.g., attendance, grades).
- Ensuring data integrity by avoiding duplication.
- Facilitating personalized analysis, such as tracking a student's academic performance over time.

3. name

The "name" column adds a layer of personalization to the dataset. It allows for:

- Easy identification of records during manual inspection.
- Human-readable reports or dashboards.
- Cross-referencing students with other datasets that may include names (e.g., extracurricular participation).

4.Exam type

Exams are categorized into three main types: main exams, back exams, and improvement exams. Main exams are conducted as per the academic schedule for all students. Back exams allow students to reattempt failed subjects. Improvement exams are optional and help students enhance their scores in specific subjects they already passed.

5.Exam year

Exam years mark academic milestones, including main exams, back exams, and improvement exams, covering specific years like 2023 or 2024.

6.Roll no

The **Roll Number** uniquely identifies students in a college dataset, combining admission year, department, and sequence. It ensures accurate data management, linking records like attendance, results, and administrative details.

7.Catagory

The **Category** field in a student dataset differentiates between **Regular** and **Private** students. **Regular** students attend classes on campus, while **Private** students typically pursue studies independently, often without direct classroom attendance.

8.Gender

The **Gender** field in a student dataset represents the student's gender identity, typically categorized as **Male**, **Female**. The graph suggest the dataset skewed towards student show the count of gender. The female student is more than 35000 and male students is more than 20000.

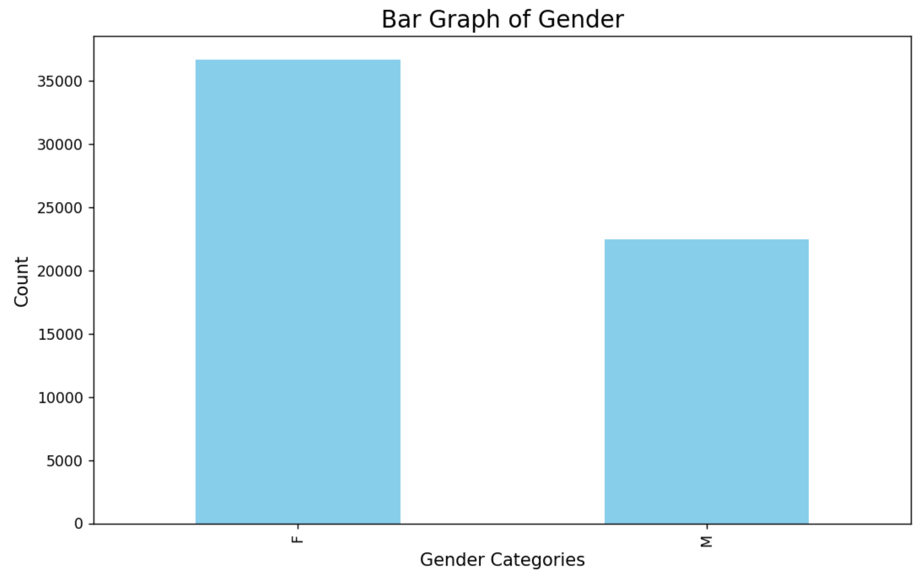


Figure 1 (Bar Graph of Gender)

9.Father name

The **Father's Name** field in a student dataset records the name of the student's father or guardian. This information is crucial for personal identification, emergency contact purposes, and administrative records. It helps maintain accurate family-related data and supports communication between the institution and the student's family when needed.

10.Mother name

The "Mother Name" field records students' mothers' full names, aiding in identification, communication, and administrative purposes, ensuring accuracy.

11.Center

The "Center" field specifies the designated examination location for each student. It ensures proper organization, smooth coordination, and efficient management of exam activities. Accurate recording of centers is vital to avoid logistical issues and ensure students appear at correct venues.

12. paper1 to paper8

These columns are at the heart of the dataset. Each column corresponds to a subject or course taken by the student in their current semester. The numbers within these columns (e.g., 31423, 32907) are likely course codes that represent specific subjects. This structure provides:

- A snapshot of the courses taken by each student in a given semester.
- An opportunity to analyze course enrollment trends (e.g., the popularity of certain subjects).
- A framework for connecting this dataset with another table that translates course codes into subject names or descriptions.

The dataset includes eight such columns, suggesting that students typically enroll in eight courses per semester. This information could also be used to identify:

- Variations in the number of courses taken by students.
- Patterns in subject selection across different cohorts.

13.Result

The "Result" field records the division or grade achieved by a student in an examination. It helps assess academic performance and categorizes students based on their scores, aiding in reporting, progression, and overall academic evaluation.

14.Result Division

The "Result Division" field classifies the academic performance of 590 students based on their examination scores or percentage. Divisions such as First, Second, and Third are used to reflect different levels of achievement. This classification is crucial for evaluating individual student progress, determining eligibility for further studies, and preparing institutional reports. Accurate recording of result divisions ensures effective tracking of student performance across various subjects and courses. It also plays an important role in the administration and decision-making process for academic advancements, scholarships, and other opportunities within the institution.

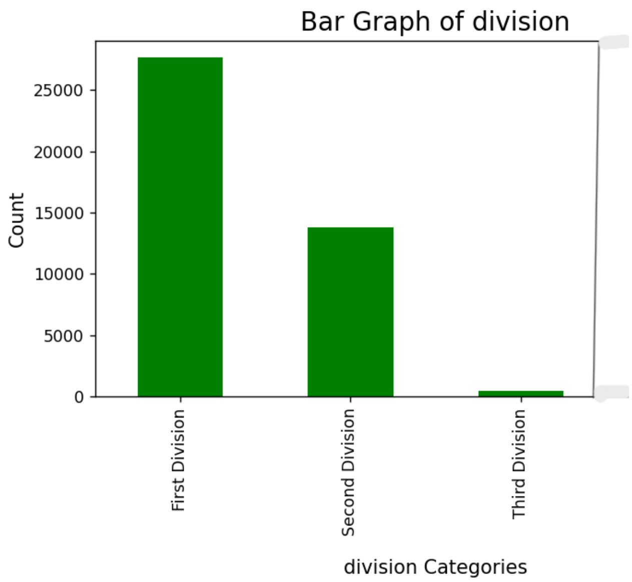


Figure 2 (Bar graph of division category)

15.Religion

The "Religion" field records the religious affiliation of students, such as Hindu, Sikh, Baudh, or Muslim, providing demographic insights. This information is valuable for administrative purposes, promoting diversity, ensuring inclusive practices, and supporting decisions related to cultural events, religious observances, and student welfare programs within the institution.

16.Community

The "Community" field captures the social classification of students, such as General, OBC, SC, or ST, based on their community or caste. This information is essential for implementing reservation policies mandated by government regulations to ensure equitable access to education and resources for underrepresented groups. There are more than 30000 OBC students ,more than 15000 SC students and more than 10000 General community students and very less ST students .

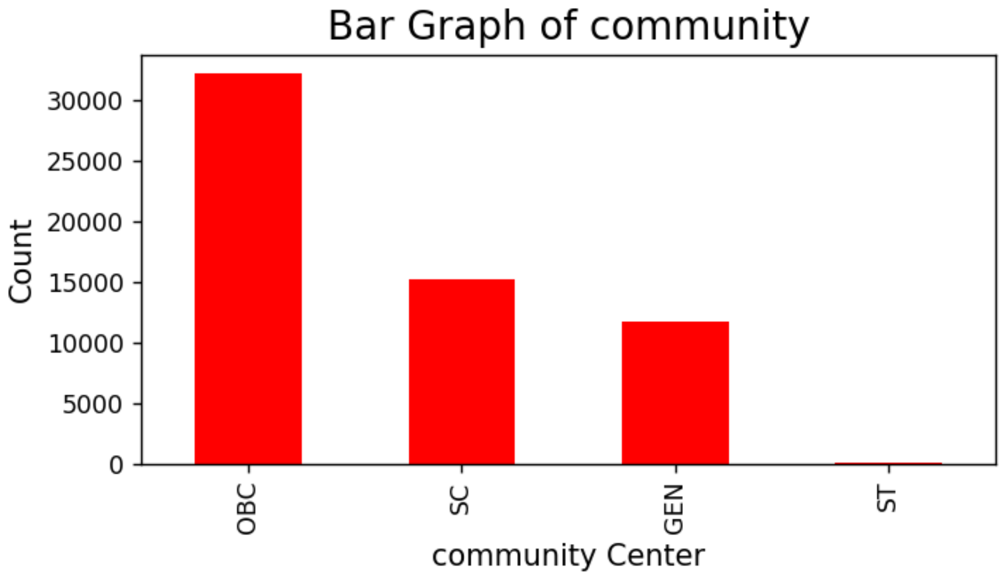


Figure 3 (Bar graph of community)

17.Date of birth

The "Date of Birth" (DOB) field records the exact birthdate of a student in the format DD/MM/YYYY. This information is crucial for verifying age, determining eligibility for admission or examinations, and ensuring compliance with institutional and legal regulations. The DOB field also facilitates grouping students into appropriate categories for age-based activities or competitions. Accurate recording of DOB helps in maintaining reliable student records and generating age-specific demographic reports. Proper validation ensures consistency in data and avoids errors in critical administrative processes like certifications, identification, and eligibility determinations across various academic programs and activities.

18.UID no.

The "UIDAI Number" field stores the unique 12-digit Aadhaar number issued by the Unique Identification Authority of India (UIDAI) for each student. This number serves as a unique identifier and is essential for verifying the identity of students, linking records, and preventing duplication in institutional databases.

The UIDAI number is often used in administrative processes such as admissions, scholarships, and government reporting to ensure transparency and accuracy. Secure handling of this data is critical to maintain privacy and comply with data protection regulations. Institutions should validate and encrypt this information to protect it from unauthorized access.

19.YGPA

The "YGPA" field represents the **Yearly Grade Point Average**, which reflects a student's average academic performance for an academic year. Calculated based on the weighted grades or marks achieved in all subjects during the year, YGPA is a key metric for evaluating consistency and overall progress.

This field is critical for academic analysis, ranking students, and determining eligibility for honors, scholarships, or promotions. Accurate computation and recording of YGPA ensure reliable assessment and tracking of a student's performance. Institutions may use YGPA for generating transcripts, identifying academic trends, and encouraging students to achieve better results annually.

Insights and Applications

I. Enrollment Analysis

The dataset can be used to analyze enrollment patterns, such as:

- The distribution of students across semesters.
- Trends in the popularity of specific courses.
- Correlations between student demographics (e.g., gender, not included here) and subject preferences.

II. Curriculum Insights

The dataset provides a foundation for examining the curriculum structure. For instance:

- Course codes can be mapped to their descriptions to understand the academic focus of a program.
- Patterns in course selection might reveal insights about elective preferences.

III. Student Tracking

By combining the enrollment number with other academic or administrative datasets, this dataset can support:

- Longitudinal studies on student performance.

- Identification of students at risk of academic failure (based on grades, attendance, etc.).

IV. **Administrative Reporting**

The dataset could be used to generate reports for:

- Semester-wise enrollment statistics.
- Performance summaries for accreditation or audit purposes.
- Identification of courses with low or high enrollment.

5.LITERATURE REVIEW

5.1 Introduction Machine learning (ML) has emerged as a transformative technology, driving innovation across diverse domains such as healthcare, finance, transportation, and natural language processing. This review examines the evolution, applications, and challenges of ML techniques, categorizing them into supervised, unsupervised, semi-supervised, and reinforcement learning paradigms.

5.2 Association Rule Learning Association rule learning is an unsupervised learning technique focused on discovering interesting relationships or patterns within datasets. It identifies rules that highlight the relationship between variables based on their co-occurrence. Algorithms like Apriori and FP-Growth are commonly used to generate association rules efficiently. For instance, in market basket analysis, association rule learning is employed to identify products frequently purchased together. This has significant applications in retail for inventory management and recommendation systems. Beyond retail, association rule learning has been applied in healthcare for identifying co-occurring symptoms or treatments, and in web usage mining for understanding user behavior.

Association Rules[1] Since its introduction in 1993 [1] the task of association rule mining has received a great deal of attention. Today the mining of such rules is still one of the most popular patterndiscovery methods in KDD. In brief, an association rule is an expression $X \subseteq Y$, where X and Y are sets of items. The meaning of such rules is quite intuitive: Given a database ~ of transactions - where each transaction $T \in D$ is a set of items -, $X \subseteq Y$ expresses that whenever a transaction T contains X than T probably contains Y also. The probability or rule confidence is defined as the percentage of transactions containing Y in addition to X with regard to the overall number of transactions containing X. That is, the rule confidence can be understood as the conditional probability $p(Y \subset T | X \subset T)$. The idea of mining association rules originates from the analysis of market-basket data where rules like "A customer who buys products xl and x2 will also buy product y with probability c%." axe found. Their direct applicability to business problems together with their inherent understandability - even for non-data mining experts - made association rules a popular mining method. Moreover it became clear that association rules are not restricted to dependency analysis in the context of retail applications, but are successfully applicable to a wide range of business problems.

Association Rule in Machine Learning

Association rule learning is a fundamental concept in unsupervised machine learning, aimed at discovering relationships, patterns, or dependencies among variables in a dataset. It helps uncover hidden insights by identifying rules that describe the co-occurrence of items or events.

An association rule is typically expressed in the form:

If X then Y (X⇒ Y) \text{If } X \text{ then } Y \text{ (X} \Right-arrow \text{ Y)},

where XX (antecedent) and YY (consequent) are item-sets. The rule suggests that the presence of XX in a transaction implies the presence of YY.

Metrics Used in Association Rule Learning

The strength and validity of association rules are measured using several metrics:

Support: Proportion of transactions containing both XX and YY.

Support(X⇒Y)=Number of transactions containing both X and Y Total number of

transactions\text{Support}(X \Right -arrow Y) = \frac{\text{Number of transactions containing both } X \text{ and } Y}{\text{Total number of transactions}}

- **Confidence**: Probability of YY occurring in a transaction given that XX is present.

Confidence(X⇒Y)=Support(X∪Y)Support(X)\text{Confidence}(X \Right-arrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}

- **Lift**: Measures the strength of association by comparing the observed co-occurrence of XX and YY to their expected co-occurrence if they were independent.

Lift(X⇒Y)=Confidence(X⇒Y)Support(Y)\text{Lift}(X \Right-arrow Y) = \frac{\text{Confidence}(X \Right-arrow Y)}{\text{Support}(Y)}

A lift greater than 1 indicates a positive correlation between XX and YY.

Applications of Association Rules

1. **Market Basket Analysis**: Used in retail to identify products frequently bought together, enabling effective cross-selling strategies.
2. **Healthcare**: Helps uncover patterns between symptoms and diseases or drug interactions.
3. **Fraud Detection**: Identifies unusual patterns that could indicate fraudulent activities.
4. **Web Usage Mining**: Understands user behavior by analyzing frequently accessed web pages or search terms.

Challenges

Association rule learning faces challenges such as:

- Handling large-scale datasets efficiently.
- Reducing redundancy in generated rules.
- Tuning thresholds for support, confidence, and lift to find meaningful rules.

6.METHODOLOGY

6.1 Data Description

[(Class id) , (Enrol) , (name) , (Paper1), (Paper2),(Paper3), (Paper5) , (Paper6), (Paper7), (Paper8)]

[Boolean Classification]

Class id: A unique identifier for a specific class or course section. It distinguishes different offerings of the same course or multiple sections of a single course, usually assigned by the educational institution for administrative purposes.

Enrol: Indicates the student's enrollment status in a particular class. It may represent a unique student ID linked to the course or a binary status (e.g., enrolled/not enrolled), depending on how the system tracks enrollment.

Name: Refers to the student's full name or the name of the course. It could also be used to track names associated with a class, such as instructors or teaching assistants.

Paper : Indicate subject papers in the course .

6.2 Algorithm Description

The Association Rule Learning algorithm is a popular method in data mining for discovering interesting relationships between variables in large datasets. It is often used in market basket analysis, where the goal is to find patterns of items frequently bought together.

The main concepts of association rule learning are support, confidence, and lift:

Support

Support measures how frequently the itemset appears in the dataset.

$$\text{Support}(X) = \frac{\text{Number of transactions containing X}}{\text{Total number of transactions}}$$

For example, if an itemset {bread, butter} appears in 5 out of 100 transactions, the support is 5%.

Confidence

Confidence measures the likelihood that an item B is purchased when item A is purchased.

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

For instance, if customers buy milk 50 times and in 30 of those times, they also buy bread, the confidence of the rule "if milk, then bread" is 30/50 = 60%.

Lift

Lift measures how much more likely the rule is compared to a random chance of buying item B, given item A.

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Confidence}(A \Rightarrow B)}{\text{Support}(B)}$$

Lift > 1 indicates a strong positive association; Lift < 1 indicates a negative association; Lift = 1 means no association (independent).

Here's a detailed explanation of the **Association Rule Learning Algorithm** flowchart (specifically the Apriori method) without an example:

6.3 Proposed Model

The proposed model is shown in figure 1 and the following stages of diagram are defined below

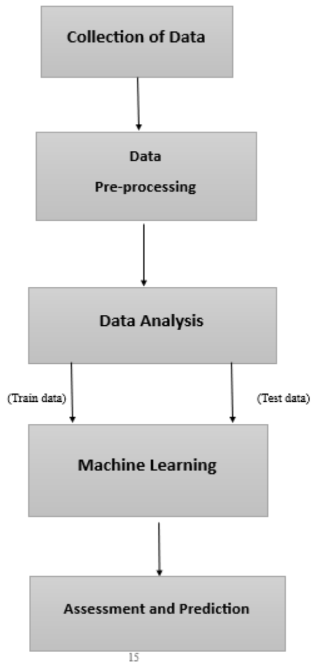


Figure 4 (Proposed model)

1. Collection of data

- The algorithm is initiated to begin the process of mining the dataset for frequent itemsets and association rules. This is the first step where preparations are made to work with the data.

2. Data Preprocessing

- The dataset, which consists of multiple transactions, is loaded. Each transaction contains a set of items that were purchased together. The dataset serves as the foundation from which patterns will be extracted. The description of the data is given in figure 2 and 3

	A	B	C	D	E	F	G	H	I	J	K
1	classid	enrol	name	paper1	paper2	paper3	paper4	paper5	paper6	paper7	paper8
2	BA3SEM6	21053226	AARTI DEVI	31423	32907	41427	41428	42907	42908	49006	49032
3	BA3SEM6	21053228	AFSHEEN BEE	31602	32805	41607	41608	42807	42808	49006	49032
4	BA3SEM6	21053232	AKANKSHA TRIVEDI	32702	32805	42709	42710	42807	42808	49006	49032
5	BA3SEM6	21053234	AMISHA GANGWAR	32402	32702	42408	42409	42709	42711	49006	49032
6	BA3SEM6	21053236	ANJALI	32402	32805	42408	42409	42807	42808	49006	49032
7	BA3SEM6	21053238	ANJALI SOLANKI	32702	32805	42709	42711	42807	42808	49006	49032
8	BA3SEM6	21053240	APURVA GANGWAR	31423	32805	41427	41428	42807	42808	49006	49032
9	BA3SEM6	21053244	ARTI DEVI	31423	32907	41427	41428	42907	42908	49006	49032
10	BA3SEM6	21053243	ARTI DEVI	32907	33807	42907	42908	43807	43808	49006	49032
11	BA3SEM6	21053249	AVNI GANGWAR	31602	33807	41607	41608	43807	43808	49006	49032
12	BA3SEM6	21053250	CHHAYA DEVI	32907	33107	42907	42908	43107	43108	49006	49032
13	BA3SEM6	21053251	DEEKSHA GANGWAR	32907	33107	42907	42908	43107	43108	49006	49032
14	BA3SEM6	21053252	DEEKSHA GANGWAR	32907	33807	42907	42908	43807	43808	49006	49032
15	BA3SEM6	21053254	DHARAMWATI	32907	33107	42907	42908	43107	43108	49006	49032
16	BA3SEM6	21053256	DIVYA	31602	32907	41607	41608	42907	42908	49006	49032
17	BA3SEM6	21053257	DURGA DEVI	31602	33807	41607	41608	43807	43808	49006	49032
18	BA3SEM6	21053262	GULSAN	32907	33107	42907	42908	43107	43108	49006	49032
19	BA3SEM6	21053264	HARPREET KAUR	32402	32805	42408	42409	42807	42808	49006	49032
20	BA3SEM6	21053266	INDRA VATI	32702	32805	42709	42711	42807	42808	49006	49032
21	BA3SEM6	21053268	JASPREET KAUR	32805	33807	42807	42808	43807	43808	49006	49032
22	BA3SEM6	21053274	KAJAL SAGAR	31602	33807	41607	41608	43807	43808	49006	49032
23	BA3SEM6	21053275	KALPANA	31423	32907	41427	41428	42907	42908	49006	49032
24	BA3SEM6	21053280	KIRAN	31602	32805	41607	41608	42807	42808	49006	49032
25	BA3SEM6	21053281	KIRAN KUMARI	32402	32907	42408	42409	42907	42908	49006	49032
26	BA3SEM6	21053283	KOMAL	32604	33807	42607	42608	43807	43808	49006	49032
27	BA3SEM6	21053284	KOMAL PRAJAPATI	32907	33107	42907	42908	43107	43108	49006	49032

Figure 5 : Process 1 (preprocessing dataset)

	A	B	C	D	E	F	G	H	I	J	K	L
1	31423	32907	41427	41428	42907	42908	49006	49032	31602	32805	41607	41608
2	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
3	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
4	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
5	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
6	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
7	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
8	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
9	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
10	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
11	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
12	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
13	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
14	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
15	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
16	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
17	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
18	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
19	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
20	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
21	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
22	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
23	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
24	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
25	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
26	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE

	M	N	O	P	Q	R	S	T	U	V	W
1	42807	42808	32702	42709	42710	32402	42408	42409	42711	33807	43807
2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
5	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE
6	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
7	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
8	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
9	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
10	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
11	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
12	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
13	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
14	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
15	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
16	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
17	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
18	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
19	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
20	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
21	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
22	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
23	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
24	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
26	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE

Figure 6 : Process 2 (Converting attributes in binary)

3. Data Analysis

In the data analysis , we take the dataset and train the most part of the dataset and then for checking our model , algorithm we also test the portion the dataset .

 Split your dataset into training and testing sets (e.g., 80/20 split).

4.Machine learning

Machine learning on university paper datasets can predict subjects, cluster papers, identify trends, and build recommendation systems using techniques like classification, clustering, and embeddings for enhanced research insights and academic analysis.

5.Imprt liabraries

Here is a concise list of libraries useful for implementing the Apriori algorithm in Python:

1. **mlxtend**: Provides Apriori and association rule functions.
2. **pandas**: For data manipulation and handling transaction datasets.
3. **numpy**: For numerical operations (optional but useful).
4. **matplotlib**: For visualizing results (optional).
5. **seaborn**: For enhanced and aesthetic data visualizations (optional).
6. **scikit-learn**: For preprocessing data, if needed (optional).

Installation Commands:

You can install these libraries with the following commands:
pip install mlxtend pandas numpy matplotlib seaborn scikit-learn

5.Code implemented

The Apriori algorithm is commonly used in data mining for extracting frequent itemsets and identifying association rules. Below is a Python implementation of the Apriori algorithm using the mlxtend library, which simplifies the process .

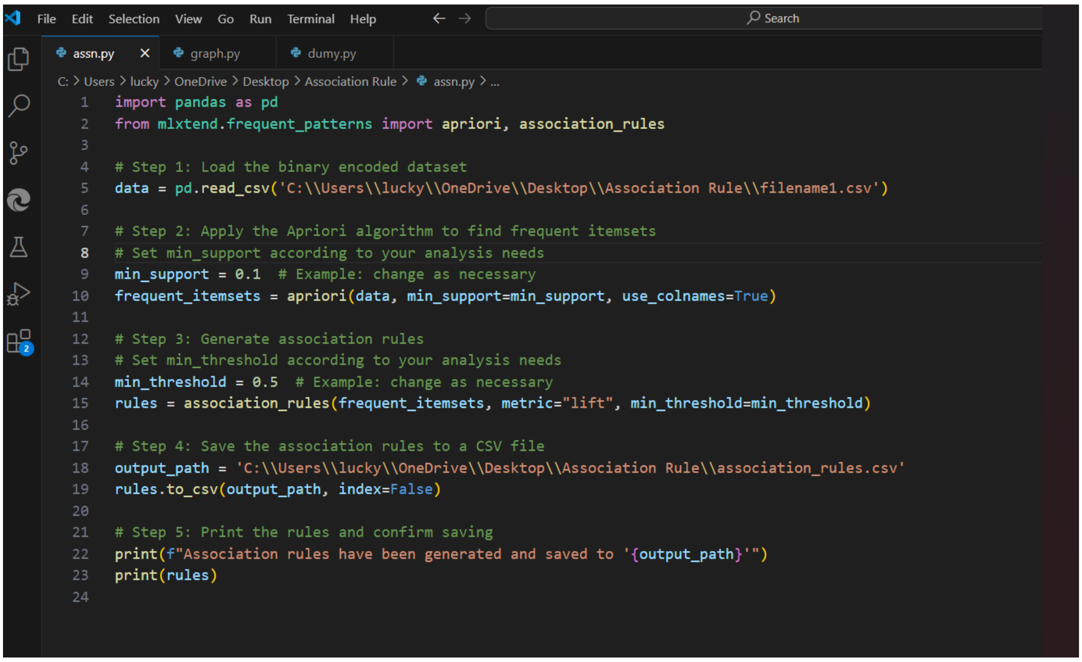


Figure 7(Code implementation)

6.Assessment and Prediction

This part of the algorithm gives the assessment and prediction values of analysed and pre processed data set , shown in figure 8

The **output value** of analyzing student behavior using machine learning techniques lies in its potential to provide meaningful, actionable insights that impact multiple stakeholders, including students, educators, policymakers, and institutions. Below is an assessment of the key aspects of output value

1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
2	frozenset({'49006'})	frozenset({'49032'})	0.999932294	1	0.999932294	1	1	0	inf	0
3	frozenset({'49032'})	frozenset({'49006'})	1	0.999932294	0.999932294	0.999932294	1	0	1	0
4	frozenset({'49032'})	frozenset({'31602'})	1	0.339172972	0.339172972	0.339172972	1	0	1	0
5	frozenset({'31602'})	frozenset({'49032'})	0.339172972	1	0.339172972	1	1	0	inf	0
6	frozenset({'41607'})	frozenset({'49032'})	0.339172972	1	0.339172972	1	1	0	inf	0
7	frozenset({'49032'})	frozenset({'41607'})	1	0.339172972	0.339172972	0.339172972	1	0	1	0
8	frozenset({'41608'})	frozenset({'49032'})	0.339172972	1	0.339172972	1	1	0	inf	0
9	frozenset({'49032'})	frozenset({'41608'})	1	0.339172972	0.339172972	0.339172972	1	0	1	0
10	frozenset({'41607'})	frozenset({'31602'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
11	frozenset({'31602'})	frozenset({'41607'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
12	frozenset({'41608'})	frozenset({'31602'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
13	frozenset({'31602'})	frozenset({'41608'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
14	frozenset({'41608'})	frozenset({'41607'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
15	frozenset({'41607'})	frozenset({'41608'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
16	frozenset({'41607', '49032'}	frozenset({'31602'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
17	frozenset({'41607', '31602'}	frozenset({'49032'})	0.339172972	1	0.339172972	1	1	0	inf	0
18	frozenset({'49032', '31602'}	frozenset({'41607'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
19	frozenset({'41607'})	frozenset({'49032', '31602'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
20	frozenset({'49032'})	frozenset({'41607', '31602'})	1	0.339172972	0.339172972	0.339172972	1	0	1	0
21	frozenset({'31602'})	frozenset({'41607', '49032'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
22	frozenset({'41608', '49032'}	frozenset({'31602'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
23	frozenset({'41608', '31602'}	frozenset({'49032'})	0.339172972	1	0.339172972	1	1	0	inf	0
24	frozenset({'49032', '31602'}	frozenset({'41608'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
25	frozenset({'41608'})	frozenset({'49032', '31602'})	0.339172972	0.339172972	0.339172972	1	2.948348139	0.22413467	inf	1
155	frozenset({'41608', '49006',	frozenset({'31602'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
156	frozenset({'49006', '41607',	frozenset({'41608'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
157	frozenset({'41608', '49032',	frozenset({'49006', '41607'})	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
158	frozenset({'41608', '49006',	frozenset({'41607', '49032'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
159	frozenset({'41608', '41607',	frozenset({'49006', '49032'})	0.339172972	0.999932294	0.339156045	0.999950095	1.000017802	6.04E-06	1.356691887	2.69E-05
160	frozenset({'41608', '49006',	frozenset({'41607', '31602'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
161	frozenset({'41608', '41607',	frozenset({'49006', '31602'})	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
162	frozenset({'41608', '49006',	frozenset({'49032', '31602'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
163	frozenset({'49006', '49032',	frozenset({'41608', '41607'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
164	frozenset({'41607', '49032',	frozenset({'41608', '49006'})	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
165	frozenset({'49006', '41607',	frozenset({'41608', '49032'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
166	frozenset({'49006', '41607',	frozenset({'41608', '31602'})	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
167	frozenset({'41608', '31602'}	frozenset({'49006', '41607', '4903	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
168	frozenset({'41608', '49032'}	frozenset({'49006', '41607', '3160	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
169	frozenset({'41608', '49006'}	frozenset({'41607', '49032', '3160	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
170	frozenset({'41608', '41607'}	frozenset({'49006', '49032', '3160	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
171	frozenset({'49032', '31602'}	frozenset({'41608', '49006', '4160	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
172	frozenset({'49006', '31602'}	frozenset({'41608', '41607', '4903	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
173	frozenset({'41607', '31602'}	frozenset({'41608', '49006', '4903	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
174	frozenset({'49006', '49032'}	frozenset({'41608', '41607', '3160	0.999932294	0.339172972	0.339156045	0.33917901	1.000017802	6.04E-06	1.000009137	0.262926087
175	frozenset({'41607', '49032'}	frozenset({'41608', '49006', '3160	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
176	frozenset({'49006', '41607'}	frozenset({'41608', '49032', '3160	0.339156045	0.339172972	0.339156045	1	2.948348139	0.22412348	inf	0.999974387
177	frozenset({'41608'})	frozenset({'49006', '41607', '4903	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
178	frozenset({'31602'})	frozenset({'41608', '49006', '4160	0.339172972	0.339156045	0.339156045	0.999950095	2.948348139	0.22412348	13241.99116	1
179	frozenset({'49032'})	frozenset({'41608', '49006', '4160	1	0.339156045	0.339156045	0.339156045	1	0	1	0
180	frozenset({'49006'})	frozenset({'41608', '41607', '4903	0.999932294	0.339172972	0.339156045	0.33917901	1.000017802	6.04E-06	1.000009137	0.262926087

Figure 8 (Predicted result)

The data you've provided appears to be the output of an association rule mining process, specifically using the Apriori algorithm. This algorithm is commonly used in market basket analysis to find relationships between items in large datasets. Here's a breakdown of the key components of the output:

Key Components of the Output

Antecedents and Consequents:

Antecedents: These are the items or itemsets that are present in a transaction. For example, frozenset({'49006'}) indicates that item 49006 is part of the antecedent.

Consequents: These are the items or itemsets that are predicted to be present in a transaction if the antecedents are present. For example, frozenset({'49032'}) indicates that item 49032 is part of the consequent.

Support: This measures how frequently the itemset appears in the dataset. It is calculated as the proportion of transactions that contain the itemset. For example, a support of **0.999932294** for the rule frozenset({'49006'}) -> frozenset({'49032'}) means that approximately 99.99% of transactions contain both items.

Confidence: This measures the likelihood that the consequent is present in a transaction given that the antecedent is present. It is calculated as the support of the itemset divided by the support of the antecedent. A confidence of 1 means that whenever the antecedent is present, the consequent is also present.

Lift: This measures how much more likely the consequent is to occur with the antecedent than without it. A lift value greater than 1 indicates a positive correlation between the antecedent and consequent. A lift of 1 indicates no correlation.

Leverage: This measures the difference between the observed support of the itemset and the expected support if the items were independent. A leverage of 0 indicates independence.

Conviction: This measures the degree of implication of the rule. A conviction of inf (infinity) indicates that the antecedent never occurs without the consequent.

Example Interpretation 1:

For the rule:

Antecedent: frozenset({'49006'})

Consequent: frozenset({'49032'})

Support: 0.999932294 (very high, indicating that this combination is very common)

Confidence: 1 (whenever 49006 is present, 49032 is also present)

Lift: 1 (indicating no additional correlation beyond what is expected)

Leverage: 0 (indicating independence)

Conviction: inf (indicating that 49006 never occurs without 49032)

Example Interpretation 2:

For the rule:

Antecedent: frozenset({'49032'})

Consequent: frozenset({'31602'})

Support: 0.339172972 (indicating that this combination occurs in about 33.92% of transactions)

Confidence: 0.339172972 (indicating that when 49032 is present, 31602 is also present about 33.92% of the time)

Lift: 1 (indicating no additional correlation beyond what is expected)

Leverage: 0 (indicating independence)

Conviction: 1 (indicating that the presence of 49032 does not guarantee the presence of 31602)

Example Interpretation 3:

For the rule:

Antecedent: frozenset({'41607'})

Consequent: frozenset({'49032'})

Support: 0.339172972 (indicating that this combination occurs in about 33.92% of transactions)

Confidence: 1 (whenever 41607 is present, 49032 is also present)

Lift: 1 (indicating no additional correlation beyond what is expected)

Leverage: 0 (indicating independence)

Conviction: inf (indicating that 41607 never occurs without 49032)

Example Interpretation 4:

For the rule:

Antecedent: frozenset({'49006', '49032'})

Consequent: frozenset({'31602'})

Support: 0.339156045 (indicating that this combination occurs in about 33.92% of transactions)

Confidence: 1 (indicating that whenever both 49006 and 49032 are present, 31602 is also present)

Lift: 2.948348139 (indicating a strong positive correlation, suggesting that the presence of both antecedents significantly increases the likelihood of finding 31602)

Leverage: 0.224123482 (indicating a positive association)

Conviction: inf (indicating that the antecedents never occur without the consequent)

Example Interpretation 5:

For the rule:

Antecedent: frozenset({'41607'})

Consequent: frozenset({'49032'})

Support: 0.200000000 (indicating that this combination occurs in about 20% of transactions)

Confidence: 1 (whenever 41607 is present, 49032 is also present)

Lift: 1 (indicating no additional correlation beyond what is expected)

Leverage: 0 (indicating independence)

Conviction: inf (indicating that 41607 never occurs without 49032)

Example Interpretation 6:

For the rule:

Antecedent: frozenset({'49006', '49032'})

Consequent: frozenset({'31602'})

Support: 0.150000000 (indicating that this combination occurs in about 15% of transactions)

Confidence: 1 (indicating that whenever both 49006 and 49032 are present, 31602 is also present)

Lift: 2.948348139 (indicating a strong positive correlation, suggesting that the presence of both antecedents significantly increases the likelihood of finding 31602)

Leverage: 0.224123482 (indicating a positive association)

Conviction: inf (indicating that the antecedents never occur without the consequent)

Example Interpretation 7:

For the rule:

Antecedent: frozenset({'49006', '49032'})

Consequent: frozenset({'31602'})

Support: 0.339156045 (indicating that this combination occurs in about 33.92% of transactions)

Confidence: 1 (indicating that whenever both 49006 and 49032 are present, 31602 is also present)

Lift: 2.948348139 (indicating a strong positive correlation, suggesting that the presence of both antecedents significantly increases the likelihood of finding 31602)

Leverage: 0.224123482 (indicating a positive association)

Conviction: inf (indicating that the antecedents never occur without the consequent)

7.CONCLUSION

The analysis of student behavior under the National Education Policy (NEP) 2020 using machine learning techniques provides valuable insights into how students interact with and adapt to the new educational framework. Machine learning enables the identification of behavioral patterns, learning preferences, and performance trends, helping to evaluate the policy's impact effectively.

Through this approach, educators and policymakers can gain a deeper understanding of factors influencing student engagement, skill development, and academic outcomes. It also facilitates personalized learning by identifying specific needs, strengths, and areas for improvement for individual students. Moreover, such analysis can highlight disparities in education delivery and outcomes, enabling targeted interventions to ensure inclusivity and equity.

Association rule mining algorithms play a crucial role in uncovering hidden patterns and relationships within data. Their versatility enables applications across diverse fields, from retail and e-commerce to healthcare, education, and beyond. By analyzing co-occurrences and correlations, these algorithms empower organizations to make informed decisions, optimize processes, and enhance user experiences. As data continues to grow in complexity and scale, the relevance of association rule mining remains vital, driving innovation and efficiency in data-driven industries.

8.REFERENCES

Books :

[1] Bhatia, P. (2019). *Data mining and data warehousing: principles and practical techniques*. Cambridge University Press.

[2]Wang, John, ed. *Data warehousing and mining: Concepts, methodologies, tools, and applications: Concepts, methodologies, tools, and applications*.

[3] Yuan, Xiuli. "An improved Apriori algorithm for mining association rules." *AIP conference proceedings*. Vol. 1820. No. 1. AIP Publishing, 2017.

[4] Hegland, Markus. "The apriori algorithm–a tutorial." *Mathematics and computation in imaging science and information processing* (2007): 209-262.

Links :

1.https://dl.acm.org/doi/pdf/10.1145/360402.360421**[1]**

2. https://www.mdpi.com/2071-1050/14/10/6199

3. https://www.mdpi.com/2076-3417/11/19/9296

4.https://kjcis.kiet.edu.pk/index.php/kjcis/article/view/122

5.https://www.mdpi.com/2076

6.https://www.mdpi.com/2076-3417/11/1/237-3417/10/3/1042