IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12820

An overview of Data Mining Technique To Find Out Student DropOut Ratio For College

Dr Himanshu Maniar

Assistant Professor, Computer, Bhagwan Mahavir College of Computer Application, Surat, India

Abstract: Predicting student dropout using data mining is an important application in the field of education. Dropout prediction models can help educational institutions identify at-risk students early and provide appropriate interventions to improve retention rates.

Keywords: Data Mining, Decision Tree, Deep Learning, Survival Analysis

I. INTRODUCTION

Student dropout rates can be influenced by a variety of factors, and these factors can vary depending on the educational level (e.g., high school, college) and the specific context (e.g., region or country).

II. COMMON REASONS FOR STUDENT DROPOUT

1. Academic Challenges: Poor academic performance or low grades can lead to frustration and a lack of motivation, eventually causing students to drop out.

2. Financial Issues: Financial constraints, including the inability to afford tuition, textbooks, or other educational expenses, can force students to leave school.

3. Family Responsibilities: Some students may have family responsibilities, such as caring for siblings or working to support their families, which can make it difficult to continue their education.

4. Lack of Support: Inadequate support from teachers, counsellors, or parents/guardians can contribute to students feeling disconnected or unsupported in their educational journey.

5. Health Problems: Physical or mental health issues can interfere with a student's ability to attend classes regularly or perform well academically.

6. School Environment: A hostile or unsafe school environment, including bullying or violence, can make students feel unsafe and cause them to drop out.

7. Lack of Relevance: If students perceive that what they are learning is not relevant to their future goals or that their education does not align with their interests, they may lose motivation to continue.

8. Peer Pressure: Negative peer influences, such as involvement in risky behaviors or peer pressure to skip school, can lead to dropout.

9. Cultural and Language Barriers: Students from diverse cultural backgrounds or those with limited proficiency in the language of instruction may face additional challenges that contribute to dropout rates.

10. Transitions: Transition points in education, such as the transition from middle school to high school or from high school to college, can be particularly vulnerable times for students to drop out.

11. Teen Pregnancy and Parenthood: Teenagers who become parents may find it challenging to balance their parental responsibilities with school, leading to dropout.

12. Substance Abuse: Substance abuse issues, including drug or alcohol addiction, can disrupt a student's education and lead to dropout.

13. Discrimination and Bias: Discrimination based on race, gender, or other factors can create an unwelcoming environment that causes some students to leave school.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12820

14. Lack of Access to Resources: Students in underserved communities may have limited access to quality educational resources, extracurricular activities, or college preparation programs.

15. Lack of Clear Educational Goals: Students who lack clear educational or career goals may struggle to see the purpose of continuing their education and may drop out.

III. COLLEGE STUDENT DROPOUT RATES

College dropout rates refer to the percentage of students who enroll in a college or university but do not complete their degree programs. These rates can vary based on various factors, including the type of institution, students' socioeconomic backgrounds, their academic preparedness, and personal circumstances. Here is more information about college dropout rates:

1. **Types of Institutions:** Dropout rates can differ significantly between different types of colleges and universities. For example:

- **Community Colleges**: Two-year community colleges tend to have higher dropout rates compared to four-year institutions. Many students attend community colleges for vocational training or to complete lower-division coursework before transferring to a four-year school.
- **Four-Year Institutions:** Four-year colleges and universities generally have lower dropout rates, particularly for students pursuing bachelor's degrees.
- **Online and For-Profit Institutions:** Dropout rates may be higher for students enrolled in online programs and for-profit institutions due to various factors, including the flexibility of online courses and financial pressures.

2. Factors Influencing College Dropout Rates:

College dropout rates can be influenced by a range of factors, including:

- **Financial Constraints:** High tuition costs, student loan debt, and the need to work to support oneself or one's family can lead some students to drop out.
- **Academic Preparedness:** Students who are not academically prepared for the rigors of college coursework may struggle and eventually drop out.
- **Lack of Support:** Insufficient academic and personal support, such as tutoring, counseling, and mentoring, can contribute to student attrition.
- **Family and Personal Circumstances:** Life events like illness, family responsibilities, or personal crises can force students to leave college temporarily or permanently.
- **Motivation and Goal Clarity:** Students who are unsure of their educational and career goals may be more likely to drop out.

3. Demographic Differences:

College dropout rates can vary among demographic groups. Factors such as race, ethnicity, gender, and age can play a role in these differences. For example, some minority groups may face unique challenges, and non-traditional students (e.g., older adults returning to college) may have different dropout patterns.

4. Policy and Intervention:

Many colleges and universities have implemented policies and interventions aimed at reducing dropout rates. These may include academic advising, financial aid programs, mentoring initiatives, and support services for underrepresented or at-risk students.

5. Data and Reporting:

Dropout rates are typically tracked and reported by educational institutions and government agencies. The calculation of dropout rates may vary, with some institutions reporting rates for a single academic year, while others track over several years.

6. Economic Implications:

College dropout rates can have economic consequences for individuals and society as a whole. Individuals who do not complete college may have lower earning potential and limited access to certain career opportunities. Additionally, high college dropout rates can be a concern for policymakers aiming to increase the educational attainment of their population.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12820

Reducing college dropout rates is a priority for many educational institutions and governments. Research on this topic often explores the specific challenges and interventions that can help students persist in their educational journeys and complete their degree programs successfully.

IV. DROPOUT PREVENTION STRATEGIES

Reducing dropout rates is a crucial goal for educational institutions and policymakers. Various dropout prevention strategies have been developed and implemented to support students and increase their chances of completing their education successfully. Here are some effective strategies for preventing dropout:

1. Early Warning Systems: Implement early warning systems that identify students who may be at risk of dropping out. These systems often use academic, attendance, and behavioural indicators to flag students who may need additional support.

2. High-Quality Counselling and Advising: Provide students with access to trained counsellors and academic advisors who can offer guidance, support, and assistance in setting and achieving educational goals.

3. Mentoring and Tutoring Programs: Establish mentoring programs where experienced students or adults can mentor at-risk students, providing them with guidance, encouragement, and a positive role model.
Offer tutoring services to help students who are struggling academically catch up on their coursework.

4. College and Career Readiness Programs: Implement programs that help students explore career options, set educational and career goals, and understand the pathways to achieving those goals. Provide information about college admissions, financial aid, scholarships, and other resources to make higher education more accessible.

5. Personalized Learning Plans: Develop personalized learning plans for students that take into account their strengths, weaknesses, interests, and goals. These plans can help students stay motivated and engaged in their education.

6. Alternative Education Options: Create alternative education pathways, such as online courses, night classes, or vocational programs, to accommodate students with unique schedules or needs.

7. Financial Support: Offer financial aid, scholarships, and grants to help students afford the costs of education, reducing the financial barriers that can lead to dropout.

8. Family Engagement: Involve parents and guardians in their children's education by providing them with information about their child's progress, encouraging participation in school activities, and offering resources to support their involvement.

9. School Climate and Safety: Foster a positive school climate that is inclusive, safe, and welcoming to all students. Address issues such as bullying and violence, as these can contribute to dropout rates.

10. Targeted Interventions: - Identify and provide targeted interventions for students who are already displaying signs of disengagement or academic struggle. These interventions can include additional academic support, counseling, or access to special programs.

11. Transition Support: - Provide support for students during critical transition points, such as the transition from middle school to high school or from high school to college. These transitions can be challenging for students and may increase the risk of dropout.

12. Data Analysis and Monitoring: - Continuously collect and analyze data on student performance, attendance, and behaviour to identify trends and patterns that can inform targeted intervention strategies.

13. Collaboration: - Foster collaboration among teachers, administrators, counsellors and other stakeholders to ensure a coordinated and comprehensive approach to dropout prevention.

14. Evaluation and Continuous Improvement: - Regularly evaluate the effectiveness of dropout prevention strategies and make adjustments based on data and feedback.

15. Community Partnerships: - Partner with community organizations, nonprofits, and businesses to provide additional support and resources to students and their families.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023
DOI: 10.17148/IJARCCE.2023.12820

Successful dropout prevention often involves a combination of these strategies tailored to the specific needs and challenges of the student population. By addressing the underlying factors that contribute to dropout and providing a supportive and engaging learning environment, educational institutions can improve student retention and success.

V. VARIOUS DATA MINING TECHNIQUES

There are various machine learning models and techniques that can be used for dropout prediction in education. Here are some common models and approaches:

1. **Logistic Regression:** Logistic regression is a straightforward and interpretable model often used for binary classification tasks like dropout prediction. It estimates the probability of a student dropping out based on input features.
2. **Decision Trees and Random Forests:** Decision trees and random forests are popular for their interpretability and ability to handle both categorical and numerical data. They can capture non-linear relationships between features and dropout.
3. **Support Vector Machines (SVM):** SVMs can be used for binary classification tasks, including dropout prediction. They work well when there is a clear margin of separation between dropout and non-dropout cases.
4. **Neural Networks:** Deep learning models, such as neural networks, can capture complex patterns in data. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are especially useful when dealing with sequential data, like student behaviour over time.
5. **Naive Bayes:** Naive Bayes classifiers are suitable when dealing with text or categorical data. They are often used in natural language processing tasks related to education, such as analysing student essays or feedback.
6. **Gradient Boosting Machines:** Gradient Boosting models like XGBoost or Light are ensemble methods that can handle complex relationships in the data. They are known for their high predictive accuracy.
7. **K-Nearest Neighbours (KNN):** KNN is a non-parametric method that predicts the class of an instance based on the majority class among its k-nearest neighbours. It can work well for small to medium-sized datasets.
8. **Time-Series Analysis:** Time-series models, such as ARIMA (Autoregressive Integrated Moving Average), can be useful when dealing with dropout prediction over time, considering seasonality and trends.
9. **Ensemble Methods:** Combining multiple models into an ensemble can often lead to improved predictive performance. Stacking, bagging, or boosting techniques can be applied to various base models.
10. **Deep Learning with Embeddings:** In the context of educational data, deep learning models with embeddings can be used to represent categorical variables like courses or student IDs in a continuous space. These embeddings can then be used as inputs to neural networks.
11. **Survival Analysis:** Survival analysis techniques, such as Cox Proportional-Hazards Models, are used when predicting dropout duration rather than binary dropout vs. non-dropout. This can be valuable for understanding the time-to-event aspect of dropout.
12. **Hybrid Models:** Combining multiple models, including traditional statistical methods and machine learning algorithms, can lead to robust dropout prediction models.

It's essential to choose the most appropriate model or combination of models based on the nature of your data, the complexity of the problem, and the interpretability requirements. The choice may also depend on the specific goals and constraints of your dropout prediction project.

VI. DECISION TREE

Decision trees are a powerful data mining technique that can be used to address college dropout ratios. Decision tree models can help identify the key factors and decision points that contribute to student attrition.

let's create a simplified example of how a decision tree can be used to analyze and address college dropout rates. In this hypothetical example, we'll consider several factors that may influence student dropout: academic performance, attendance, socioeconomic status, and involvement in extracurricular activities.

Step 1: Data Collection and Preprocessing

Imagine we have collected data on 200 college students, and for each student, we have the following information:

Academic Performance (GPA): High, Medium, Low
Attendance: Good, Moderate, Poor
Socioeconomic Status: High, Medium, Low

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

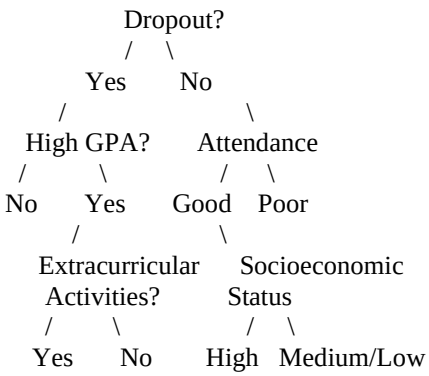
DOI: 10.17148/IJARCCE.2023.12820

Extracurricular Activities: Yes, No
Dropout Status: Yes (dropped out) or No (did not drop out)

Step 2: Building the Decision Tree

Let's build a decision tree using this data to predict the likelihood of a student dropping out based on the factors mentioned above.

Here's a simplified decision tree:



Step 3: Interpretation

High GPA: If a student has a high GPA, they are less likely to drop out (follow the "No" branch).
Low GPA: If a student has a low GPA, we further consider their attendance.
Good Attendance: If attendance is good, the student is less likely to drop out (follow the "No" branch).
Poor Attendance: If attendance is poor, we consider their socioeconomic status.
High Socioeconomic Status: If socioeconomic status is high, the student is less likely to drop out (follow the "No" branch).
Medium/Low Socioeconomic Status: If socioeconomic status is medium or low, the student is more likely to drop out (follow the "Yes" branch).
Extracurricular Activities: If a student is involved in extracurricular activities, they are less likely to drop out (follow the "No" branch). However, this factor is considered only if the student has a low GPA and poor attendance.

Step 4: Intervention Strategies

Based on the decision tree analysis, we can derive some intervention strategies:
Identify students with low GPAs and poor attendance as high-risk.
Provide additional academic support and counselling for these high-risk students.
Encourage involvement in extracurricular activities as a potential dropout prevention strategy, particularly for high-risk students.

Step 5: Continuous Monitoring and Improvement

Continuously collect data on students and their outcomes, update the decision tree model as needed, and refine intervention strategies based on ongoing analysis. This iterative process can lead to improved student retention and success.

VII. CONCLUSION

In conclusion, addressing the college dropout ratio is a multifaceted challenge that requires a comprehensive and data-driven approach. It is a critical concern for educational institutions, policymakers, and society as a whole, as it not only impacts individual students but also has economic and social implications. Here are some key takeaways and strategies for solving the college dropout ratio:

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023
DOI: 10.17148/IJARCCE.2023.12820

❖ **Data-Driven Decision-Making:** Utilizing data mining techniques, predictive analytics, and machine learning models can help colleges identify at-risk students and develop targeted interventions.
❖ **Early Intervention:** Identifying students who may be at risk of dropping out early in their educational journey allows for timely and effective interventions. Early warning systems and continuous monitoring are essential.
❖ **Personalized Support:** Recognizing that each student has unique needs and challenges, colleges should provide personalized support, including academic advising, mentoring, tutoring, and counseling.
❖ **Holistic Approach:** Addressing factors contributing to dropout, such as academic performance, attendance, socioeconomic status, and personal circumstances, requires a holistic approach that considers both academic and non-academic aspects of a student's life.
❖ **Transitions and Milestones:** Special attention should be given to critical transition points, such as the transition from high school to college or from lower-division to upper-division coursework, as these are times when students may be more vulnerable to dropping out.
❖ **Inclusivity and Equity:** Efforts to reduce dropout rates should address disparities in access and outcomes among different demographic groups, ensuring that all students have equal opportunities for success.
❖ **Ethical Considerations:** Data privacy, informed consent, and ethical use of student data are paramount. Institutions must adhere to legal and ethical guidelines when implementing data-driven strategies.
❖ **Continuous Improvement:** Dropout prevention strategies should be continually evaluated and refined based on feedback, data analysis, and evolving best practices.
❖ **Community Engagement:** Collaboration with community organizations, employers, and other stakeholders can enhance support networks and resources available to students.
❖ **Communication and Transparency:** Clear and transparent communication with students, families, and the wider community is vital for building trust and ensuring that students are aware of available support services.

Solving the college dropout ratio is an ongoing process that requires a commitment to student success, innovation, and adaptability. By harnessing the power of data and employing a holistic, student-centred approach, colleges and universities can increase student retention and improve the educational experience for all. Ultimately, addressing the dropout challenge contributes to a more educated and skilled workforce and a stronger society.

REFERENCES

[1]. Cortez, P., & Silva, A. M. (2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito & J. Teixeira (Eds.), Proceedings of the 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008), pp. 5-12.
[2]. Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.
[3]. Kotsiantis, S. B., & Pierrakeas, C. (2003). Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. In Proceedings of the 5th Panhellenic Conference with International Participation, Volos, Greece, pp. 1-8.
[4]. Sezer, C., Dogantekin, E., & Dogantekin, A. (2013). A Comparative Study of Classification Algorithms in Terms of an Educational Data Mining Application. Computers in Human Behavior, 29(4), 1592-1598.
[5]. Baker, R. S., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1(1), 3-17.

Books:

[1]. Baker, R. S. (2019). Educational Data Mining: An Advance for Intelligent Systems in Education (2nd ed.). CRC Press.
[2]. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.
[3]. Baker, R. S., O'Neil, D. K., & Beck, J. E. (Eds.). (2019). Educational Data Mining: Applications and Trends. Springer.
[4]. Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme Learning Machine: Theory and Applications. Neurocomputing, 70(1-3), 489-501.
[5]. Duan, K., Keerthi, S. S., & Poo, A. N. (2003). Evaluation of Simple Performance Measures for Tuning SVM Hyperparameters. Neurocomputing, 51, 41-59.

