MINOR PROJECT LITERATURE SURVEY

Project titled **"Scholar Lens: Analysis of Student Learning Data for Early Intervention"** focuses on using data mining and machine learning (ML) to predict students at risk of academic decline and to suggest early interventions. This aligns with **Sustainable Development Goal (SDG-4): Quality Education**.

Suggested Literature Areas:

1. **Early Warning Systems in Education**:
    - Research papers discussing ML models for predicting at-risk students.
    - Case studies of schools or universities implementing predictive analytics.
2. **Data Mining in Education**:
    - Approaches to extracting actionable insights from student data like grades, attendance, and online activity.
    - Studies on the impact of such insights on intervention strategies.
3. **ML Algorithms for Academic Prediction**:
    - Comparative studies on algorithms like Random Forests, SVM, Neural Networks, etc., used for educational data.
4. **Educational Data Representation**:
    - Techniques for representing learning data, including feature engineering (e.g., time spent on platforms, interaction levels).
5. **Impact of Interventions**:
    - Effectiveness of timely interventions based on predictive analytics.

1. **EARLY WARNING SYSTEMS IN EDUCATION**

    a. **Research papers discussing ML models for predicting at-risk students.**

Paper 1 :

Paper : Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models

Authors : MUHAMMAD ADNAN , ASAD HABIB , JAWAD ASHRAF , SHAFAQ MUSSADIQ , ARSALAN ALI RAZA , MUHAMMAD ABID , MARYAM BASHIR , AND SANA ULLAH KHAN

Publisher : IEEE

Abstract : Online learning platforms like MOOCs, VLEs, and LMS offer flexible education to large audiences but face challenges such as high dropout rates, low engagement, and students' lack of self-regulation. This study proposes a predictive model to identify at-risk students early, enabling timely interventions by instructors to improve engagement and performance. Various ML and DL algorithms were evaluated using metrics such as accuracy, precision, recall, and F-score. Random Forest (RF) was found to deliver the best performance, with accuracy, precision, and recall improving progressively at different course lengths. Key predictors included assessment scores, engagement intensity (clickstream data), and time-dependent variables. This model empowers instructors to address student dropouts proactively.

Link : https://ieeexplore.ieee.org/abstract/document/9314000

Paper 2 :

Paper : Course Success Prediction and Early Identification of At-Risk Students Using Explainable Artificial Intelligence

Authors : Berat Ujkani , Daniela Minkovska and Nikolay Hinov

Publisher : MDPI : electronis

Abstract : This study leverages Artificial Intelligence (AI) to predict student success on online education platforms using the Open University Learning Analytics Dataset (OULAD). Employing machine learning, deep learning, and SHapley Additive exPlanations (SHAP) as an Explainable AI (XAI) technique, it identifies key factors influencing success or failure, such as student engagement and registration timelines. Unlike traditional methods, this AI-driven approach provides interpretable insights into student performance, enabling targeted interventions for at-risk students. Achieving up to 94% accuracy, the findings offer practical guidance for educators, administrators, and policymakers to enhance online learning effectiveness.

Link : https://www.mdpi.com/2079-9292/13/21/4157

b. Case studies of schools or universities implementing predictive analytics.

Paper 1 :

Paper : Predictive Analytics Machinery for STEM Student Success Studies

Authors : Lingjun He , Richard A. Levine , Andrew J. Bohonakb , Juanjuan Fanc ,and Jeanne Stronach

Publisher : Taylor & Francis Group

Abstract : This study develops a predictive analytics framework to enhance STEM student success by leveraging random forest machine learning algorithms. The framework provides automated, timely insights for assessing student performance, informing pedagogical decisions, and identifying at-risk students. Key features include identifying input thresholds, quantifying factors influencing success, and generating student success scores. Applied at San Diego State University, the framework integrates statistical analyses with business intelligence tools to design, validate, and deploy dashboards, improving STEM persistence and graduation outcomes. This approach empowers educators and advisers with actionable insights to optimize student support strategies.

Link : https://www.tandfonline.com/doi/full/10.1080/08839514.2018.1483121

Paper 2 :

Paper : Predictive Models for Educational Purposes: A Systematic Review

Authors : Ahlam Almalawi , Ben Soh , Alice Li and Halima Samra

Publisher : MDPI : big data and cognitive computing

Abstract : This systematic review examines predictive models in education, focusing on their use in forecasting student performance, identifying at-risk students, and personalizing learning. Machine learning (ML) algorithms, including SVMs, ANNs, and Decision Trees, consistently outperform traditional models in handling complex, non-linear educational data and improving predictive accuracy. While ML models effectively integrate socio-economic, demographic, and academic data, challenges include biases in historical data, lack of transparency, and difficulties in interpreting AI decisions. The study emphasizes the need for equitable, interpretable models, standardized data collection, and ethical handling of student data to enhance trust and generalizability in AI-driven educational tools.

Link : https://www.mdpi.com/2504-2289/8/12/187

2. **Data Mining in Education**:

    a. **Approaches to extracting actionable insights from student data like grades, attendance, and online activity.**

Paper 1 :

Paper : Analysing the Impact of Online Learning on Higher Education: A Text Analytics Approach

Authors : Gulam Ruti Asplangyi

Publisher : International Journal of Information Technology and Computer Science Applications (IJITCSA)

Abstract : The ongoing Covid-19 pandemic has necessitated a significant shift in higher education towards online learning modalities, which has raised concerns regarding student learning outcomes and academic performance. Institutions like Bloom have observed a decline in average grades and are responding by leveraging text analytics to analyze unstructured data from various sources, including social media and educational platforms. This approach aims to identify patterns and insights that can inform strategic interventions to enhance academic performance and mitigate the adverse effects of the pandemic. The article outlines Bloom's framework for implementing text analytics, addressing challenges, proposing solutions, and detailing implementation strategies to optimize educational outcomes during these turbulent times.

Link : https://ejurnal.jejaringppm.org/index.php/jitcsa/article/view/149

Paper 2 :

Paper : Insight and Action Analytics: Three Case Studies to Consider

Authors : Mark David Milliron, Ph.D. Civitas Learning Laura Malcolm, M.A. Civitas Learning David Kil, M.S.E.E., M.B.A. Civitas Learning

Publisher : Research & Practice in Assessment

Abstract : Civitas Learning fosters a community of practice among higher education leaders to leverage insight and action analytics for student success. By integrating data from multiple institutional sources (e.g., ERP, LMS, CRM), predictive models are developed to enhance student progression and completion. This article examines three case studies demonstrating how tailored approaches address unique institutional challenges, leading to significant positive outcomes. A key insight is that no universal predictive model fits all institutions. The article concludes with key findings and considerations for future work in data-driven student success initiatives.

Link : https://eric.ed.gov/?id=EJ1062814

b. Studies on the impact of such insights on intervention strategies

Paper 1 :

Paper : An examination of the efficacy of INSIGHTS in enhancing the academic and behavioural development of children in early grades.

Authors : O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G.

Publisher : APA PsycNet

Abstract : This group randomized trial examined the effectiveness of *INSIGHTS Into Children's Temperament* in enhancing academic achievement, sustained attention, and reducing disruptive behavior in low-income kindergarten and 1st-grade students. Twenty-two urban schools were randomly assigned to either INSIGHTS or a supplemental reading program as a control condition. Data from 435 students across 122 classrooms were analyzed using hierarchical linear models. Results showed that children in INSIGHTS had greater growth in math and reading achievement, improved sustained attention, and fewer behavior problems compared to the control group. Behavioral improvements partially mediated academic gains. Findings highlight the benefits of social-emotional interventions in supporting early academic success.

Link : https://psycnet.apa.org/doi/10.1037/a0036615

Paper 2 :

Paper : An Evaluation of the Effects of "INSIGHTS" on the Behavior of Inner City Primary School Children

Authors : Sandee McClowry, David L. Snow, and Catherine S. Tamis-LeMonda

Publisher : Springer Nature Link

Abstract : This study evaluated the impact of INSIGHTS Into Children's Temperament, a temperament-based intervention, on reducing behavior problems in inner-city primary school children. A randomized trial compared INSIGHTS to a Read Aloud control program in six schools. Participants included 148 first- and second-grade children, their parents, and 46 teachers. Using parental reports, results showed that INSIGHTS significantly reduced children's behavior problems at home compared to the control group, particularly among children diagnosed with disruptive disorders. Findings suggest that temperament-based interventions can effectively improve behavioral outcomes in at-risk children, emphasizing the need for early preventive programs in urban schools.

Link : https://link.springer.com/content/pdf/10.1007/s10935-005-0015-7.pdf

3. **ML Algorithms for Academic Prediction**:
    i. **Comparative studies on algorithms like Random Forests, SVM, Neural Networks, etc., used for educational data.**

Paper 1 :

Paper : Comparison of support vector machine, random forest and neural networkclassifiers for tree species classification on airborne hyperspectral APEXimages

Authors : Edwin Raczko and Bogdan Zagajewski

Publisher : Taylor & Francis Group

Abstract : This study compares three machine learning algorithms (Support Vector Machines, Random Forest, and Artificial Neural Networks) for classifying five common tree species using hyperspectral data from the Airborne Prism Experiment sensor. Using a 0.632 bootstrap evaluation procedure, Artificial Neural Networks achieved the highest median classification accuracy at 77%, followed by Support Vector Machines (68%) and Random Forest (62%). While ANN showed superior accuracy, RF and SVM demonstrated more stable results with lower variance in their performance metrics.

Link : https://www.tandfonline.com/doi/epdf/10.1080/22797254.2017.1299557?needAccess=true

Paper 2 :

Paper : Comparative Study of Supervised Algorithms for Prediction of Students' Performance

Authors : Madhuri T. Sathe , Amol C. Adamuthe

Publisher : I.J. Modern Education and Computer Science

Abstract : This study evaluates various machine learning algorithms for predicting student academic performance across three different educational datasets (school, college, and e-learning platform). The research compares C5.0, J48, CART, Naïve Bayes, K-Nearest Neighbor, Random Forest, Support Vector Machine, and Artificial Neural Network classifiers with different parameter tunings. Results demonstrate that Random Forest and C5.0 algorithms consistently outperform other methods in predicting student academic performance across all three datasets.

Link : https://www.researchgate.net/profile/Amol-Adamuthe/publication/349712082_Comparative_Study_of_Supervised_Algorithms_for_Prediction_of_Students'_Performance/links/66f154388f4e4465a3c36093/Comparative-Study-of-Supervised-Algorithms-for-Prediction-of-Students-Performance.pdf

4. **Educational Data Representation**:
    i. **Techniques for representing learning data, including feature engineering (e.g., time spent on platforms, interaction levels).**

Paper 1 :

Paper : Design and user behaviour analysis of an English learning social platform based on digital entertainment content recommendation algorithm

Authors : Wenhua Liu

Publisher : ELSEVIER

Abstract : This research explores enhancing distance education through a social network-based English learning platform. The study implements a collaborative filtering algorithm to analyse learners' historical data and create personalized learning recommendations by matching students with similar learning behaviours and preferences. The platform incorporates social features to facilitate student interaction and knowledge sharing. Experimental results and user feedback demonstrate that this approach significantly improves both learning outcomes and inter-student communication compared to traditional English learning platforms.

Link : https://www.sciencedirect.com/science/article/pii/S1875952124001022

Paper 2 :

Paper : Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning

Authors : Ángel F. Agudo-Peregrina a , Santiago Iglesias-Pradas a , Miguel Ángel Conde-González b , Ángel Hernández-García

Publisher : ELSEVIER

Abstract : This study examines the relationship between student interactions in virtual learning environments and academic performance across two learning modalities: online and VLE-supported face-to-face courses. The research develops three system-independent classifications of interactions and analyzes their impact using data from eight courses (six online, two face-to-face). Results reveal that certain types of interactions significantly correlate with academic performance in online courses, while showing no significant relationship in VLE-supported face-to-face courses. This suggests that the importance of digital interactions varies based on the learning modality.

Link : https://www.sciencedirect.com/science/article/pii/S074756321300188X

5. **Impact of Interventions**:

    a. **Effectiveness of timely interventions based on predictive analytics.**

Paper 1 :

Paper : Using Machine Learning to Predict and Improve Student Learning Outcomes in Real Time

Authors : Md Salman , V.V.S.S.S. Balaram

Publisher : International Journal of Web of Multidisciplinary Studies

Abstract : This study explores the application of machine learning algorithms in educational settings to predict and enhance student learning outcomes. Through experimental analysis of ensemble methods and neural networks, we demonstrate how real-time data analytics can enable personalized interventions and improve academic achievement. Our findings establish a framework for implementing machine learning in education and highlight its potential for creating adaptive learning environments.

Link : http://ijwos.com/index.php/home/article/view/3

Paper 2 :

Paper : Improving student success using predictive models and data visualisations

Authors : Alfred Essa and Hanan Ayad

Publisher : Association for Learning Technology

Abstract : This paper introduces the Student Success System (S3), a comprehensive analytical platform designed to address global educational challenges, particularly high dropout rates and low graduation rates. S3 employs machine learning and statistical methods to identify at-risk students early, providing data visualization tools and case management capabilities for effective interventions. The system's open architecture enables seamless integration with existing software, creating an end-to-end solution for improving student retention and academic success through predictive analytics.

Link : https://journal.alt.ac.uk/index.php/rlt/article/view/1359