*Conference Name: International Conference on Science & Technology, 18-19 July 2024, Bali*
*Conference Dates: 18-Jul- 2024 to 19-Jul- 2024*
*Conference Venue: Ibis Bali Kuta, Jl. Raya Kuta No. 77, 80361 Kuta, Bali, Indonesia*
*Appears in: MATTER: International Journal of Science and Technology (ISSN 2454-5880)*
*Publication year: 2024*

*Jain et. al., 2024*

*Volume 2024, pp. 57-58*

*DOI- https://doi.org/10.20319/icstr.2024.5758.*

*This paper can be cited as: Jain, H., Ajumobi, K., White, K., Sugumaran, V., Rotheberger, M. (2024). Predicting college dropout likelihood based on High School and College Data: A Machine Learning Approach. International Conference on Science & Technology, 18-19 July 2024, Bali. Proceedings of Social Science and Humanities Research Association (SSHRA), 2024, 57-58.*

# PREDICTING COLLEGE DROPOUT LIKELIHOOD BASED ON HIGH SCHOOL AND COLLEGE DATA: A MACHINE LEARNING APPROACH

**Hemant Jain**
*Data Analytics Department, University of Tennessee Chattanooga, Chattanooga, TN*
*hemant-jain@utc.edu*

**Korede Ajumobi**
*TN*

**Keith White**
*Public Education Foundation, Chattanooga, TN*

**Vijayan Sugumaran**
*Oakland University, Rochester, MI*

**Marcus Rotheberger**
*University of Nevada Las Vegas, Las Vegas, NV*

## Abstract

*College dropout rate is a significant problem, especially in the US higher education system. Among all undergraduate students, up to 40% drop out before completing their degree. This significantly impacts students and Universities alike financially and in wasted efforts. Previous research shows that there are early indicators of college success in the high school record such as grades, attendance, disciplinary incidents, and ACT/SAT scores. Additionally, there are factors in college*

*experiences, especially in the first year of college which impact the likelihood of dropout. However, to the best of our knowledge, there is no comprehensive model that can accurately predict the likelihood of college dropout and provide an early warning either in high school and/or the first year of college. We were fortunate to get access to longitudinal ten-year data of high school graduates of public schools in a county in the United States and were able to follow a subset of students who went to a specific public University. Based on more than one hundred variables from high school and college records and students' final status we trained various machine learning models to predict the likelihood of student dropout and identify factors that play a significant role. Based on this information a prototype decision support system was developed and evaluated.*

**Keywords:**

Machine Learning, Student Success in College, Dropout Rate, Prediction