Structural Neural Networks Meet Piecewise Exponential Models for Interpretable College Dropout Prediction

Chuan Cai
University of Delaware
Newark, DE, USA
chuanc@udel.edu

Adam Fleischhacker
University of Delaware
Newark, DE, USA
ajf@udel.edu

We propose a novel approach to address the issue of college student attrition by developing a hybrid model that combines a structural neural network with a piecewise exponential model. This hybrid model not only shows the potential to robustly identify students who are at high risk of dropout, but also provides insights into which factors are most influential in dropout prediction. To evaluate its effectiveness, we compared the predictive performance of our hybrid model with two other survival analysis models: the piecewise exponential model and a hybrid model combining a fully-connected neural network with a piecewise exponential model. Additionally, we compared it to five other cross-sectional models: Ridge Logistic Regression, Lasso Logistic Regression, CART decision tree, Random Forest, and XGBoost decision tree. Our findings demonstrate that the hybrid model outperforms or performs comparably to the other models when predicting dropout among students at the University of Delaware in Spring 2020, Spring 2021, and Spring 2022. Moreover, by categorizing predictors into three distinct groups—academic, economic, and social-demographic—we discovered that academic predictors play a prominent role in distinguishing between dropout and retained students, while other predictors may indirectly influence predictions by impacting academic variables. Consequently, we recommend implementing an intervention program aimed at identifying at-risk students based on their academic performance and activities, with additional consideration for economic and social-demographic factors in customized intervention plans.

Keywords: college dropout, structural neural network, piecewise exponential model, interpretable hybrid model

1. INTRODUCTION

College dropout continues to be a significant concern for higher education institutions (Albreiki et al., 2021; Aulck et al., 2016; Cannistra et al., 2022). Data from the National Center for Education Statistics (NCES) reveals that approximately 40% of first-time, full-time degree-seeking undergraduate students at 4-year degree-granting institutions fail to obtain a bachelor's degree within six years at the same institution. Alarmingly, around 20% of these students drop out within their first year (NCES, 2022). The consequences of dropout are substantial, leading to wasted resources for students, institutions, and society as a whole. Students who discontinue their college education not only waste their time but also the tuition and fees they have paid and the loans they have borrowed, and tended to have lower income and an increased risk of living in poverty (Bouchrika et al., 2023). Concurrently, institutions suffer losses in terms of resources dedicated to these students, as well as potential tuition revenue and future alumni donations. On average, institutions lose approximately $10 million in tuition revenue annually due to attrition (Raisman, 2013). Furthermore, both federal and state governments waste their appropriations to institutions and grants to students, with an estimated expenditure of $9 billion between 2003 and 2008 on students who withdrew within their first year (Schneider, 2010).

Two main approaches have been developed to address the issue of college dropout: theory-driven and data-driven approaches (Cannistra et al., 2022). The theory-driven approach focuses on constructing conceptual models to comprehend the underlying reasons for dropout. It considers students' decisions to discontinue their education as an interplay of various factors, including family background, demographic characteristics, academic performance, social integration, organizational determinants, personal satisfaction, and institutional commitment (Bean, 1980; Spady, 1970; Tinto, 1975). By developing these conceptual models and analyzing observed data, specialized recommendations can be derived to mitigate attrition (Bean, 1980). In contrast, the data-driven approach emphasizes dropout prediction. Statistical and machine learning models are utilized to forecast students' academic performance and/or identify those who are at risk of dropping out (Aulck et al., 2016; Baker & Yacef, 2009; Heredia-Jimenez et al., 2020). However, this approach often involves a trade-off between interpretability and predictability. Recent studies have proposed various methods to improve the interpretability of predictive models (Agarwal et al., 2021; Baranyi et al., 2020; Cohausz et al., 2022; Kopper et al., 2021).

Our study aims to integrate the strengths of both approaches, as educational institutions require both explanatory insights and predictive capabilities to develop effective intervention plans for reducing attrition (Wagner et al., 2023). We predict the dropout risks of students who are currently enrolled at the start of an academic semester, more specifically, whether they will fail to enroll for the subsequent semester. This allows the student advising team sufficient time to develop and implement effective intervention strategies throughout the semester and assists the Budget Office to estimate the tuition revenue from the current students. To model college students' dropout risks, we employ a piecewise exponential model (Kopper et al., 2021; Friedman, 1982). This model is well-suited for analyzing longitudinal processes in students' academic careers, which involve time-varying factors and effects associated with dropout risks. While this classic survival analysis model offers interpretability, it may not provide optimal predictive performance. To enhance predictability, we introduce a neural network model into the piecewise survival analysis to capture the hazard, which transitions from a linear combination of variables in traditional survival analysis to a nonlinear function of the variables.

However, fully-connected neural network models pose challenges to interpretability due to their black-box nature. To address this dilemma, we employ a structural neural network (Fan et al., 2022; Köhler and Langer, 2021; Schmidt-Hieber, 2020), where we impose a structure inspired by theoretical frameworks of student attrition onto the neural networks. Specifically, we categorize variables into three groups: academic, economic, and socio-demographic. Variables within each category interact with each other, forming hidden layers that generate a final neuron representing the category. In total, three final neurons are generated. These final neurons are then linearly combined to predict the hazard, which is subsequently converted into dropout probabilities. Consequently, our model not only provides a list of students with a high risk of dropout but also identifies whether students are more likely to dropout due to academic performance (Stinebrickner & Stinebrickner, 2014), financial burden (Cai & Fleischhacker, 2022), or social integration (Stage, 1989).

The primary objective of this study is to address two key research questions that will contribute to the design of an effective intervention plan for attrition reduction:

1. Does the utilization of a structural neural network enhance predictive performance compared to traditional survival analysis models that employ linear hazards?

2. Which category or categories of variables impact(s) the prediction of students' dropout risks?

2. Literature Review and Conceptual Framework

Extensive academic research has been conducted on college student dropout, employing both theory-driven and data-driven approaches. The theory-driven approach aims to establish a theoretical foundation and develop conceptual models to comprehend students' dropout decisions. Tinto (1975) formulated a theoretical model based on Durkheim's suicide theory (Durkheim, 2005) and a cost-benefit analysis, which sought to explain dropout decisions through the interaction between individuals and institutions. The model suggested that factors such as family background, individual characteristics, and pre-college schooling influence individuals' integration within the academic and social systems of colleges, subsequently impacting their commitment to educational goals and institutional commitment. Lower levels of commitment were found to be associated with higher dropout probabilities. Bean (1980) developed a causal model inspired by turnover models in work organizations, positing that the interaction between students' background characteristics and organizational factors affects their satisfaction, institutional commitment, and ultimately dropout probabilities. The model was empirically tested using multiple regression and path analysis, utilizing questionnaires returned by 1,195 new freshmen. Gender-specific recommendations were provided to reduce attrition. Similarly, Spady (1970) developed a sociological model of the college dropout process, drawing from Durkheim's suicide theory (Durkheim, 2005). Spady argued that family background influences students' academic potential and normative congruence, subsequently impacting their grade performance, intellectual development, and friendship support. The interaction of these factors influences social integration, ultimately leading to dropout decisions. These conceptual models have provided theoretical guidance for subsequent empirical studies (Aina et al. 2022; Deike 2003; Márquez-Vera et al. 2016).

Our conceptual framework draws inspiration from Tinto's theoretical model (Tinto, 1975) and aims to capture the longitudinal process of dropout risk accumulation. We posit that students' dropout risks are determined by their goal commitment and institutional commitment, which are influenced by the interplay of academic integration, economic integration, and social integration. Academic integration is shaped by factors such as college cumulative GPA, the number of classes with a grade below D, the number of credits registered, and engagement in multiple majors. Pre-college schooling indicators, such as high school GPA and the number of Advanced Placement (AP) credits, also contribute to academic integration. Economic integration is influenced by variables including expected family contribution (EFC), outstanding balance, and the amount of financial aid received in the form of grants, scholarships, and loans. Social integration encompasses factors such as eligibility for Pell Grants (whether total family income is less than or equal to $50,000 U.S.) and being a first-generation college student, as well as demographic characteristics including gender and racial ethnicity. Importantly, the dropout risk of a student evolves each semester, reflecting the changing dynamics of the three integrations. These changes stem from time-varying factors like academic performance and financial aid, the evolving processes that generate the integrations, and the dynamic interactions among them.

The data-driven approach, as an alternative, emphasizes prediction, specifically the identification of students at high risk of dropout as candidates for targeted intervention and remedial programs (Quadri and Kalyankar, 2010). Various statistical and machine learning methods have been employed in this approach. Almarabeh (2017) compared five classification methods, including Naïve Bayes, Bayesian network, decision tree with ID3, decision tree with C4.5, and multilayer perceptron neural network, to predict the dropout risks of 225 students using 10 predictors. The Bayesian network model exhibited the best performance across various error measures, such as accuracy, true positive rate, false positive rate, and F-score. Similarly, Sandoval-Palis et al. (2020) compared logistic regression and neural network models to predict the dropout risks of 2,097 students in an engineering university in Ecuador, using four predictors, regime, leveling course type, application grade, and vulnerability index, where regime, leveling course type, and application grade were academic factors, and vulnerability index was derived from 25 socio-economic variables. The neural network models outperformed logistic regression models in terms of accuracy and AUC score. However, despite their high predictive power, neural network models are challenging to interpret due to their black-box nature.

Regression and decision tree models are preferred when the predictive performance is satisfactory, as their interpretability can aid institutions in establishing effective intervention policies. Wagner et al. (2023) compared three explainable methods and two ensemble methods to predict degree dropout in a middle-sized German university. The explainable methods are decision trees, K-nearest neighbors, and logistic regression, and the ensemble methods are AdaBoost and Random Forest. Among these, logistic regression was found to exhibit the best overall predictive performance. Nevertheless, the study also highlighted that these models did not equally excel in predicting student subpopulations, particularly concerning gender and specific study programs. Quadri and Kalyankar (2010) proposed a hybrid method for dropout prediction in an Indian institution, combining decision trees to identify relevant predictors, such as parents' income and previous semester's grades, with logistic regression for predicting students' dropout risks. Aulck et al. (2016) utilized regularized logistic regression, Random Forest, and K-nearest neighbors to predict the attrition of 32,538 students from the University of Washington, finding that regularized logistic regression performed the best. Strong predictors included GPA in math, English, chemistry, and psychology classes. Heredia-Jimenez et al. (2020) employed Random Forest to predict at-risk students across 65 undergraduate programs in a public engineering-oriented university in Ecuador. They obtained reliable predictions by excluding socio-demographics and pre-college entry information, relying solely on academic information as predictors. Ameri et al. (2016) compared the performance of

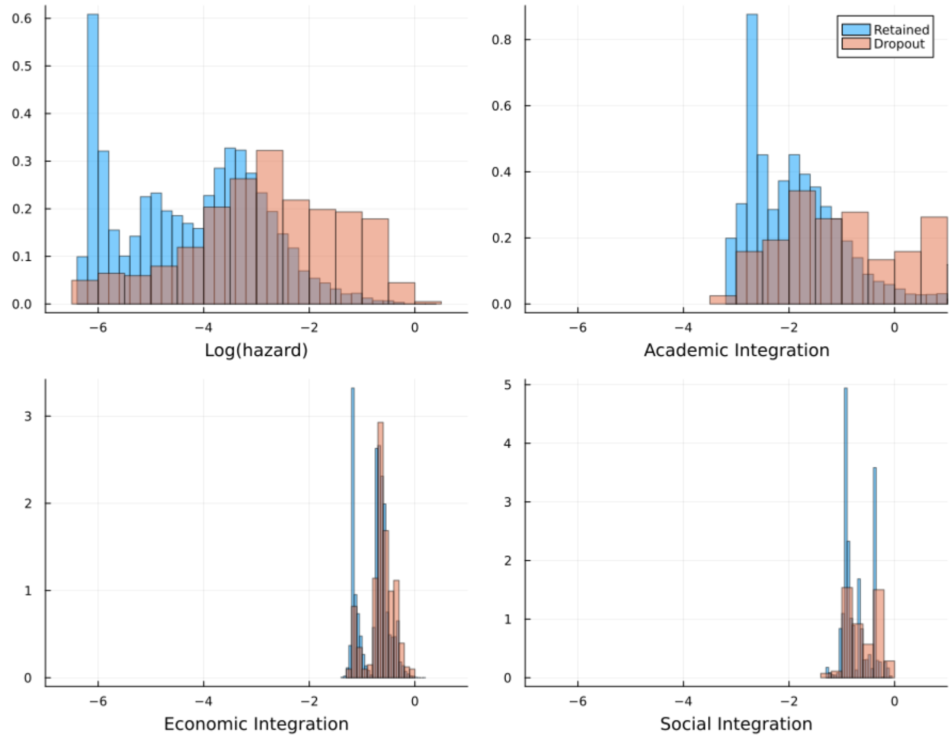

Figure 4: Histograms of log(hazard), academic integration, economic integration, and social integration in Spring 2022

5.3. Tracking At-Risk Students

In addition to predicting student dropout in the test dataset, we also apply the trained PEM-SNN models to the hold-out dataset to observe typical patterns of dropout probabilities. As explained in Section 4, the hold-out dataset comprises approximately 10% of data from all available semesters leading up to the semester of the test dataset. On average, the dropout probabilities for retained students are approximately 0.013, while dropout students exhibit an average probability of 0.101. Interestingly, many dropout students have predicted probabilities of zero for all semesters in the hold-out dataset, indicating that they initially appear likely to continue their enrollment but subsequently experience a dramatic decline in their academic performance, leading to dropout. For those dropout students with predicted probabilities higher than zero, many of them have at least one semester with a high dropout risk (dropout probability > 0.5), irrespective of whether it is their last semester or not. These findings highlight the dynamic nature of dropout risk, where students' circumstances and risks can change over time.

6. Limitations and Future Work

Our study is constrained by several limitations. First, this study is limited to a single educational institution and covers data from Fall 2012 to Spring 2022. As a result, the findings may not be directly applicable to other institutions or different time periods. Dropout risks at other institutions may be directly influenced by economic or social factors. It is essential to exercise caution when generalizing the results beyond the scope of this study. Second, we do not differentiate between students who drop out entirely from higher education and those who transfer to another institution. These two groups of students may have varying reasons for leaving their current institution. Building separate models to predict dropout risks for each group could potentially enhance predictive performance and provide more nuanced insights into the specific causes of attrition. Third, similar to the findings of Wagner et al. (2023), our models exhibit differential performance in predicting dropout risks for various student subpopulations, such as gender and Pell Grant eligibility. This suggests that the models may not be equally effective in identifying at-risk students across all demographic or socioeconomic groups, highlighting the need for further investigation and model refinement. Fourth, the imbalance in our dataset, stemming from the rarity of dropout cases, poses significant challenges to achieving more accurate predictions, as evidenced by the relatively low F1-scores.

Future research can address these limitations and explore several avenues for improvement. First: modifying the input variables and network structure of the PEM-SNN models to investigate which academic variable(s) contribute more significantly to the academic integration. This analysis can help identify the specific academic metrics that exert a more substantial influence on dropout risk. By pinpointing these metrics, institutions can focus their monitoring efforts on areas that have a higher impact on student outcomes. Second: extending the analysis to explore how economic and social factors affect the prediction of academic integration. Understanding the interactions between economic and social factors and their influence on academic integration can provide valuable insights. For instance, recognizing that first-generation college students may face unique challenges and have distinct risk factors can inform targeted interventions and support programs. Third: validating the developed models on datasets from different institutions or time periods to assess their generalizability. This external validation can help determine the extent to which the models' predictive capabilities hold across diverse educational settings. Fourth: developing tailored models for specific student subpopulations, acknowledging that different demographic or socioeconomic groups may exhibit distinct risk profiles. By customizing models to these groups, institutions can enhance their ability to identify and support at-risk students effectively.

7. Conclusion

In this study, we have introduced a hybrid model, PEM-SNN, which combines a structural neural network with a piecewise exponential model, with the goal of addressing student attrition in colleges. Our hybrid model not only shows the potential to robustly identify the students at high risk of dropout but also provides insights into the direct contributions of academic, economic, and social factors. We applied the model to predict dropout occurrences during the Spring semesters of 2020, 2021, and 2022 at the University of Delaware. The results consistently demonstrated that the PEM-SNN model exhibits the 2nd best predictive performance compared to two other PEM-related models and five cross-sectional models while providing enhanced interpretability. This indicates that the PEM-SNN model is a valuable tool for the student advising team, providing a list of at-risk students who require focused attention and support.

Among the three integrations incorporated into our model, academic integration emerged as the most influential factor in distinguishing between dropout and retained students, as evidenced by the results generated by the PEM-SNN models. Retained students tended to have lower academic integration values compared to dropout students, resulting in reduced hazard and, consequently, lower dropout probabilities. In contrast, economic and social integrations displayed similar distributions between dropout and retained students, suggesting that they do not directly contribute to dropout prediction. However, it is important to note that these economic and social factors may be very important through the indirect impact they have on academic performance. Thus, it remains important to consider a holistic approach when designing customized intervention plans for at-risk students, taking into account the interconnected nature of these factors.

In conclusion, our hybrid model, PEM-SNN, represents a valuable tool in the ongoing effort to reduce student attrition in higher education. By leveraging predictive analytics, institutions can identify at-risk students early and enabling timely interventions to enhance student success and retention. However, it is essential to continue refining and expanding these models, taking into account the unique characteristics of different institutions and student populations. Ultimately, such efforts can contribute to higher graduation rates and improved educational outcomes for all students.

Declaration of generative AI software tools in the writing process

During the preparation of this work, the authors used ChatGPT in all sections in order to improve the readability and clarity of this paper. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

AGARWAL, R., MELNICK, L., FROSST, N., ZHANG, X., LENGERICH, B., CARUANA, R. AND HINTON, G.E. 2021. Neural additive models: Interpretable machine learning with neural nets. *Advances in neural information processing systems*, *34*, 4699-4711.

AINA, C., BAICI, E., CASALONE, G. AND PASTORE, F. 2022. The determinants of university dropout: A review of the socio-economic literature. *Socio-Economic Planning Sciences*, *79*, 101102.

ALBREIKI, B., ZAKI, N., AND ALASHWAL, H. 2021. A systematic literature review of student' performance prediction using machine learning techniques. *Education Sciences 11*, 9, 552.

ALMARABEH, H. 2017. Analysis of students' performance by using different data mining classifiers. *International Journal of Modern Education and Computer Science 9*, 8, 9.

AMERI, S., FARD, M. J., CHINNAM, R. B., AND REDDY, C. K. 2016. Survival analysis based framework for early prediction of student dropouts. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. Association for Computing Machinery, New York, NY, USA, 903–912.

AULCK, L., NAMBI, D., VELAGAPUDI, N., BLUMENSTOCK, J., AND WEST, J. 2019. Mining university registrar records to predict first-year undergraduate attrition. In *Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019)*, C. F. Lynch, A. Merceron, M. Desmarais, and R. Nkambou, Eds. International Educational Data Mining Society, 9–18.

AULCK, L., VELAGAPUDI, N., BLUMENSTOCK, J., AND WEST, J. 2016. Predicting student dropout in higher education. arXiv preprint, arXiv:1606.06364.

BAKER, R. S. AND 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.

BARANYI, M., NAGY, M. AND MOLONTAY, R. 2020, October. Interpretable deep learning for university dropout prediction. In *Proceedings of the 21st annual conference on information* *technology education*. Association for Computing Machinery, New York, NY, USA, 13-19.

BEAN, J. P. 1980. Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education 12*, 155–187.

BOUCHRIKA, I. 2023. *College dropout rates: 2023 statistics by race, gender & income.* https://research.com/universities-colleges/college-dropout-rates Accessed: 12-05-2023.

CAI, C. AND FLEISCHHACKER, A. 2022. The effect of loan debt on graduation by department: A Bayesian hierarchical approach. *Journal of Student Financial Aid 51*, 2, Article 5.

CANNISTRA, M., MASCI, C., IEVA, F., AGASISTI, T., AND PAGANONI, A. M. 2022. Early-predicting dropout of university students: An application of innovative multilevel machine learning and statistical techniques. *Studies in Higher Education 47*, 9, 1935–1956.

COHAUSZ, L. 2022. Towards Real Interpretability of Student Success Prediction Combining Methods of XAI and Social Science. In *Proceedings of the 15th International Conference on Educational Data Mining (EDM 2022)*, A. Mitrovic and N. Bosch, Eds. International Educational Data Mining Society, 361–367.

COHAUSZ, L., TSCHALZEV, A., BARTELT, C. AND STUCKENSCHMIDT, H. 2023. Investigating the Importance of Demographic Features for EDM-Predictions. In *Proceedings of the 16th International Conference on Educational Data Mining (EDM 2023)*, M. Feng, T. Käser and P. Talukdar, Eds. International Educational Data Mining Society, 125–136.

DEIKE, R.C. 2003. A study of college student graduation using discrete-time survival analysis. Ph.D. thesis, The Pennsylvania State University.

DURKHEIM, E. 2005. *Suicide: A study in sociology*. Routledge.

FAN, J., KE, Z. T., LIAO, Y., AND NEUHIERL, A. 2022. Structural deep learning in conditional asset pricing. Available at SSRN: https://ssrn.com/abstract=4117882 or http://dx.doi.org/10.2139/ssrn.4117882.

FRIEDMAN, M. 1982. Piecewise exponential models for survival data with covariates. *The Annals of Statistics 10*, 1, 101–113.

HEREDIA-JIMENEZ, V., JIMENEZ, A., ORTIZ-ROJAS, M., MARÍN, J. I., MORENO-MARCOS, P. M., MUÑOZ-MERINO, P. J., AND KLOOS, C. D. 2020. An early warning dropout model in higher education degree programs: A case study in Ecuador. In *Proceedings of the Workshop on Adoption, Adaptation and Pilots of Learning Analytics in Under-represented Regions co-located with the 15th European Conference on Technology Enhanced Learning (LAUR@ EC-TEL 2020)*. P. J. M. Merino, C. D. Kloos, Y.-S. Tsai, D. Gasevic, K. Verbert, M. Perez-Sanagustin, I. Hilliger, M. A. Z. Prieto, M. Ortiz-Rojas and E. Scheihing, Eds. 58–67.

INNES, M., SABA, E., FISCHER, K., GANDHI, D., RUDILOSSO, M. C., JOY, N. M., KARMALI, T., PAL, A., AND SHAH, V. 2018. Fashionable modelling with flux. *arXiv preprint arXiv:1811.01457*.

KÖHLER, M. AND LANGER, S. 2021. On the rate of convergence of fully connected deep neural network regression estimates. *The Annals of Statistics 49*, 4, 2231–2249.

KOPPER, P., PÖLSTERL, S., WACHINGER, C., BISCHL, B., BENDER, A. AND RÜGAMER, D. 2021, May. Semi-structured deep piecewise exponential models. *In Proceedings of AAAI Spring Symposium on Survival Prediction - Algorithms, Challenges, and Applications (2021)*, R. Greiner, N. Kumar, T. A. Gerds, M. van der Schaar, Eds. PMLR, 40–53.

KOPPER, P., WIEGREBE, S., BISCHL, B., BENDER, A. AND RÜGAMER, D. 2022, May. DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis. *In Pacific-Asia Conference on Knowledge Discovery and Data Mining (2022)*. Springer International Publishing, 249–261.

LUBIN, M., DOWSON, O., GARCIA, J. D., HUCHETTE, J., LEGAT, B., AND VIELMA, J. P. 2023. JuMP 1.0: Recent improvements to a modeling language for mathematical optimization. Mathematical Programming Computation. *Mathematical Programming Computation. 15*: 581–589. https://doi.org/10.1007/s12532-023-00239-3.

MANRIQUE, R., NUNES, B.P., MARINO, O., CASANOVA, M.A. AND NURMIKKO-FULLER, T., 2019, March. An analysis of student representation, representative features and classification algorithms to predict degree dropout. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. Association for Computing Machinery, New York, NY, USA, 401-410.

MÁRQUEZ-VERA, C., CANO, A., ROMERO, C., NOAMAN, A. Y. M., MOUSA FARDOUN, H., AND VENTURA, S. 2016. Early dropout prediction using data mining: a case study with high school students. *Expert Systems*, 33: 107–124. doi: 10.1111/exsy.12135.

NATIONAL CENTER FOR EDUCATION STATISTICS (NCES). 2022. Undergraduate retention and graduation rates. https://nces.ed.gov/programs/coe/indicator/ctr Accessed: 05-22-2023

QUADRI, M. AND KALYANKAR, D. N. 2010. Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology 10*, 2, 2–5.

RAISMAN, N. 2013. The cost of college attrition at four-year colleges & universities-an analysis of 1669 US institutions. *Policy perspectives*.

SANDOVAL-PALIS, I., NARANJO, D., VIDAL, J., AND GILAR-CORBI, R. 2020. Early dropout prediction model: A case study of university leveling course students. *Sustainability 12*, 22, 9314.

SCHMIDT-HIEBER, J. 2020. Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics*, 48, 4, 1875–1897. https://doi.org/10.1214/19-AOS1875

SCHNEIDER, M. 2010. Finishing the first lap: The cost of first year student attrition in America's four-year colleges and universities. Presented *in the 50th annual Meeting of the Association for Institutional Research (AIR 2010)*.

SPADY, W. G. 1970. Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange 1*, 1, 64–85.

STAGE, F. K. 1989. Motivation, academic and social integration, and the early dropout. *American Educational Research Journal 26*, 3, 385–402.

STINEBRICKNER, R. AND STINEBRICKNER, T. 2014. Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics 32*, 3, 601–644.

TINTO, V. 1975. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research 45*, 1, 89–125.

WAGNER, K., VOLKENING, H., BASYIGIT, S., MERCERON, A., SAUER, P., AND PINKWART, N. 2023. Which approach best predicts dropouts in higher education? In *Proceedings of the*

15th International Conference on Computer Supported Education (CSEDU 2023), SCITEPRESS – Science and Technology Publications, 2, 15–26.

8. APPENDICES

8.1. APPENDIX A

Table A.1 provides an overview of the headcount flow for the ten cohorts across different semesters. The smallest cohort, Fall 2020, consisted of 3,711 first-time full-time students, while the largest cohort, Fall 2017, comprised 4,258 students. The headcount of each cohort gradually decreased over the semesters due to dropout. Blank cells in the table indicate that the corresponding cohort had not reached that particular semester.

Table A.1: Cohort headcount flow by semester

Cohort\Semester Sequence	1st	2nd	3rd	4th	5th	6th
Fall 2012	3,806	3,677	3,490	3,398	3,308	3,284
Fall 2013	3,794	3,660	3,438	3,340	3,244	3,197
Fall 2014	4,168	4,035	3,835	3,735	3,642	3,569
Fall 2015	4,092	3,944	3,725	3,644	3,538	3,497
Fall 2016	3,943	3,807	3,576	3,479	3,385	3,297
Fall 2017	4,285	4,116	3,827	3,708	3,582	3,481
Fall 2018	4,242	4,091	3,849	3,732	3,658	3,609
Fall 2019	4,136	3,934	3,705	3,621	3,554	3,484
Fall 2020	3,711	3,573	3,389	3,267	3,171	3,117
Fall 2021	4,263	4,079	3,852	3,775		

Table A.2 presents the headcounts for each semester between Fall 2012 and Spring 2022, organized by cohort. Each semester featured up to three cohorts: freshman, sophomore, and junior. Senior cohorts were excluded from the analysis as they had already completed their third year or exceeded the sixth semester. The initial semesters do not have three cohorts due to the absence of data before the Fall 2012 cohort.

Table A.2: Headcounts between Fall 2012 and Spring 2022 by cohort

Semester \ Cohort	Freshman Cohort	Sophomore Cohort	Junior Cohort	Total
Fall 2012	3,806			3,806
Spring 2013	3,677			3,677
Fall 2013	3,794	3,490		7,284
Spring 2014	3,660	3,398		7,058
Fall 2014	4,168	3,438	3,308	10,914
Spring 2015	4,035	3,340	3,284	10,659
Fall 2015	4,092	3,835	3,244	11,171
Spring 2016	3,944	3,735	3,197	10,876
Fall 2016	3,943	3,725	3,642	11,310
Spring 2017	3,807	3,644	3,569	11,020

Semester \ Cohort	Freshman Cohort	Sophomore Cohort	Junior Cohort	Total
Fall 2017	4,285	3,576	3,538	11,399
Spring 2018	4,116	3,479	3,497	11,092
Fall 2018	4,242	3,827	3,385	11,454
Spring 2019	4,091	3,708	3,297	11,096
Fall 2019	4,136	3,849	3,582	11,567
Spring 2020	3,934	3,732	3,481	11,147
Fall 2020	3,711	3,705	3,658	11,074
Spring 2021	3,573	3,621	3,609	10,803
Fall 2021	4,263	3,389	3,554	11,206
Spring 2022	4,079	3,267	3,484	10,830