Adonis Shareef

College Dropout Prediction Analysis - Case Study

- Business Understanding: Defining the Problem

  During this case study we had our team create a predictive model with the collaboration of universities including Oklahoma State, Arizona State, Iowa State, and the University of Alabama. The goal of this predictive model is to determine the potential of a student dropping out of the universities. This is important because if we can track and predict those at risk, we can help those students before they dropout.

- Data Understanding

  The data consist of features that you would expect every student to have like their GPA, their major, number of classes and their year (freshmen to senior). Also, we have track of the dropout feature with a simple yes or no response because this data is a collection from over 3 years. With the amount of information, you can have when a student registers for a program and their classes we have plenty of features that we can look into and help us make predictions on.

- Data Preparation

  Most of this data is good for modeling but we will have to change things like yes or no and true and false to integers to feed good numeric inputs to our model. There was a 5% missing data in our rows which we could simply grab the average for those student's GPA and filling those null values with the average from the rest of the data set, which will save it from skewing the data. Also, because we are across multiple universities and

span over a three-year range, we needed to also make sure there were not any duplicate rows as students might have switched schools. Now that all the data is in one data set, we must take that precaution.

- Modeling

  Our team was tasked with creating a Multiple Linear Regression model to predict the potential that a student will dropout of the university. Due to our team using this model we had to split our data into training and testing with a 70:30 split. We also had the opportunity to do use K-Neighbor's classifier methods alongside our MLR. And to remove bias using cross-validation split on our data was something we focused on.

- Deployment

  After the deployment of our model the American Public University System was able help prevent 5% of undergraduate dropouts. This was successful as a first deployment because if you think about the number of undergrad students there are in the universities, we had data on alone that would be in the hundreds of thousands maybe could hit a million students. The model is continuing to collect data as the semesters continue so we have more data and can make better predictions. The improvements that can now be made are huge because the goal is to have no student dropout but there is only so much you can do for someone but being able to see which students are at risk can really help the universities retain their students which is a win-win for both the student and the university.

- Summary and Conclusions

  Our team did great work preparing the data so that they could then build and train the model. Being able to help the students was the goal of the American Public University System and we accomplished that at a scale we can be proud of. However, there is a lot more work that can go into this model because as new courses and programs are developed there will be more thins that we see are successful and things we see that just do not work. When we start to see these trends then we can finally make better predictions and get to the core of the problems students are having. There could be features that are not as relevant that we have passed to the model that we can take out in the future to increase the accuracy of our predictions. The most important thing we can do is use these predictions to reach out to these at-risk students, so they have some support that our team made possible.

References

- Overview of Different Approaches to Deploying Machine Learning Models in Production. (2019). https://www.kdnuggets.com/2019/06/approaches-deploying-machine-learning-production.html
- School Dropout Rates in Iowa: State-wide and by Subgroup | mydata.iowa.gov. (2021). https://mydata.iowa.gov/Primary-Secondary-Ed/School-Dropout-Rates-in-Iowa-State-wide-and-by-Sub/negh-sf87/data