SAS Hackathon | Fall 2024

Track 2 | Ethical Data Analysis | Predicting Graduate School Admissions

Overview

- Welcome to iLink University! As a recently hired Predictive Modeler in our Admissions Department, you will help us decide who gets into our 2025 M.S. in Analytics Program.
- This track is great for students with a background in predictive modeling and machine learning. In other words, it is the more advanced of the two tracks offered in the 2024 SAS Hackathon.

Learning Objective

- Engage students with a relevant predictive modeling exercise and expose them to the Visual Data Mining and Machine Learning tools in SAS Model Studio and SAS Visual Analytics.
- You will also learn about the Fairness and Bias assessment tools within SAS Model Studio.

Output

- SAS Hackathon submissions are in the form of two videos submitted on your SAS Hackathon Team Page. So, complete the entire project first and then consider how you want to frame your videos. You'll have just a small amount of time to discuss the approach, the methods, and the outcomes, but it is your story to tell!

Part 1: Business as Usual

Your Task List

- You're ready to get to work, because you're a Hacker at heart. And you're provided with 1000 admission decisions from the past 5 years. So, step 1 is to better understand the general setup of the data, which has the following variables:

Variable	Label	Definition
Admitted	Admitted (Yes=1)	When Admitted = 1, the student is offered admissions into the iLink University M.S. in Analytics Program
Analytics_Work_Experience	Analytics Work Experience	Number of years working in the field of analytics.
Country_Region	Country Region	Region of the world applying from
Cultural_Identity	Cultural Identity	Cultural identity

Gender	Gender Identity or Gender at Birth	Gender identity or gender at birth
ID	Application ID	Application ID
Legacy_Admission	Legacy Admission	Legacy admission means that either (1) the student's parents attended the university or (2) they previously completed another degree at iLink University.
Mission_Statement	Mission Statement	Optional mission statement (maximum of 100 words)
Standardized_Test_Score	Standardized Test Score	Standardize test score (Z-score)
Strength_of_Recommendation	Strength of Recommendations	Overall strength of recommendations (0 to 5, higher is better)
Undergrad_Degree	Undergraduate Degree Category	Undergraduate degree category
Years_Work_Experience	Years Work Experience	Total years of work experience, all fields

- With a better understanding of the data, the next steps are ones you know by heart:
  - Get the data into SAS Viya.
  - Explore the data and get to know it A LOT better.
  - Run a bunch of predictive models using the Visual Data Mining and Machine Learning tools in SAS Viya.
  - Find the best model predicting the historical sample. Crown it champion.
- You'll then apply that champion model to a new data set and then choose the 40 students that will be admitted as part of the incoming class of 2025.
  - Please share the general characteristics of the students admitted under this approach – and anything interesting you noted as part of the modeling process.

Part 2: Business (not) as Usual

Overview

- From your studies you know that using historical data to predict new cases – particularly in the case of outcomes like admission – often perpetuates the status quo.
  - That can be a good thing when it's an equitable world, but bad when it allows bias to persist.
- Thinking that there's likely some bias in that approach, you argue to the iLink University leadership that they should statistically examine whether there are any implicit biases in their previous admission history – bias that would be carried forward without any thoughtful correction.

- Fortunately, you are very persuasive. Your Department Chair would like you to examine the models with an eye on ethical data analysis best practices.

Your (refined) Task List

- You will use the bias assessment tool in SAS Model Studio to determine if there is any historical bias in the admission process at iLink University.
- If there is any bias, you need to correct it.
  - Rerun your models while better accounting for the implicit bias, if it exists.
  - Is there a trade-off between model fit and using potentially biased variables?
- Finally share your aggregated findings
  - Compare the characteristics of those admitted in Part 1 to those admitted in Part 2
  - Comment on what changed in the models – and what you would do with more data. Please also suggest different forms of data that could be captured in future applications, if relevant.

Appendix: Learning Resources

- SAS Skill Builder for Students Courses
  - Machine Learning Using SAS® Viya®
  - Interactive Machine Learning in SAS® Viya®
  - Responsible Innovation and Trustworthy AI
- SAS Hackathon Enablement Week Recordings
  - Welcome to SAS Viya Week + Overview of SAS Viya Ecosystem
  - Data Discovery, Part 2: Predictive Modeling and Machine Learning
  - Encore, Part 1: Ethical Data Analysis