Predicting Academic Performance for Student Athletes

Fall 2024 Project Description

SPARK – CDS – DS701

Link to Notion Here

### **Client Name and Description:**

The Boston University Department of Athletics is dedicated to excellence in athletics, academics and community. Over the long history of the department our student-athletes have represented BU formidably, achieving and maintaining a high level of excellence both on and off the field while at the same time displaying commitment, integrity and sportsmanship.

* Kate Bergstrom, Senior Associate Director of Athletics, [kbergs@bu.edu](mailto:kbergs@bu.edu)
* Brittany Kane, Senior Associate Director of Athletics, [bmkane@bu.edu](mailto:bmkane@bu.edu)

### **Description:**

BU Athletics is interested in understanding how well a student will perform academically during their time at Boston University based on the student's high school GPA, geography, demographics, language proficiency for international students, what they choose to study, and standardized exam score. This will help BU Athletics understand 1) how they can better support future student athletes and 2) to assist our coaches in the recruiting process to understand if/where a student may struggle or thrive as a part of our community.

The goal is to identify how we can better support future student-athletes by finding trends in areas where students may struggle or excel within the Boston University community.

You will be provided with data from students who have already graduated, including information such as race, college, high school GPA, hometown, and semester GPA. This data needs to be analyzed to identify trends in the success or struggles of athletes. If time permits, students can also explore creating a basic regression model to make predictions based on the trained dataset.

***\*\*Please note you will not be measuring athletic performance related to admissions data and instead will focus on academic performance.\*\****

### Ideal Output & Final Deliverables:

* A cleaned and modified dataset uploaded to Google Drive and GitHub
  + Include academic performance predictions
* A final report that summarizes your findings:
  + Data analysis from previous years.
  + Visualizations illustrating key insights.
    - The deliverable for the base questions should include a minimum of 7-10 visualizations ([see here for recommendations](https://www.data-to-viz.com)) that address the base questions
  + An understanding of factors influencing student performance, such as high school GPA, geography, demographics, language proficiency for international students, and chosen field of study.
* Documentation of the prediction model and hosted on github or huggingface

| Project Details | |
| --- | --- |
| Questions | 1. What is the range of accepted SAT/ACT scores and highschool GPA for student athletes?    1. What percentage of these student athletes are domestic students? What about international students?    2. Does English being the primary language of the country impact the students' performance?    3. How does the academic performance of students with similar HS GPAs compare? (those with SAT/ACT vs. those without SAT/ACT score submitted)? Ex: A student with a 3.0 HS GPA compared academic performance in college and AI (have submitted SAT/ACT score) 2. What is the range of BU GPA for student athletes? 3. How do these students with high SAT/ACT scores or high GPA perform academically at BU compared to student athletes with low SAT/ACT scores or low GPA? 4. Does the average student athlete improve their GPA throughout their college year? 5. Do students from certain geographic areas (Northeast) perform better academically than another geographic area (Southwest)? 6. Are there any significant differences in the academic performance of student-athletes based on their sport? |
| Preferred Tech Stack | * List of tools/ tech stack preferred by client e.g. PowerBI, Flourish, etc. * Include links to recommended libraries or data cleaning solutions where relevant   + Looker Studio |
| Data Sets + Data Dictionary | Dataset:   * [Data Provided by Chen (Sept 2024)-scrubbed.xlsx](https://docs.google.com/spreadsheets/d/1sZ3JsX-xUh9NRDNFyEURmpzA6YiVm065/edit?usp=drive_link&ouid=113007906765725987232&rtpof=true&sd=true) |
|  | * Github, Trello, Google folder, Figma, etc. * Links to any relevant notes from the client |
| Project Milestones | * **Client kickoff**: Data review + questions + create slack channel * **Data Collection & Cleaning**: Receive and clean the dataset, then upload it to Google Drive. * **Exploratory Data Analysis (EDA)**: Identify trends and correlations through data visualization. * **Feature Engineering**: Create relevant features from high school GPA, geography, demographics, etc. * **Model Training**: Train and evaluate machine learning models for academic performance prediction. * **Model Tuning**: Optimize the model and validate its performance with a test set. * **Predictions**: Generate predictions for Fall 2023 and 2024 students. * **Final Report**: Summarize findings, provide visualizations, and share the final cleaned dataset, predictions, and report. * **Final Client Presentation** |
| Recommended steps | * Desk Research: Familiarize yourself with the [BU Athletics site](https://goterriers.com/index.aspx), background readings, datasets, and project description. Prepare questions for the client kickoff meeting. * Data cleaning & Machine Learning Tool Creation * Devise a data cleaning, processing, machine learning tooling, tech stack, analysis, and visualization plan. Get it signed off by the client. * Provide frequent updates to the client so they can provide feedback. |
| **Additional Details** | |
| Background Readings | Understand different types of sports analytics projects:   * <https://www.interviewquery.com/p/sports-analytics-projects> * <https://www.projectpro.io/article/sports-analytics-projects-ideas-and-examples/630> |
| Common misconceptions, and solutions roadblocks (If relevant) | * *Roadblocks/ common blockers in this project.* * *Common data cleaning challenges or tips* * *Common contextual misunderstandings* * *Tips from previous groups on this project.* |
| Ethical considerations | The predictions will NOT be used to decide on student admission. The BU athletics team doesn’t support and is not affiliated with admissions of students. |

### Contact Information:

| **Role** | **First Name** | **Last Name** | **Email** |
| --- | --- | --- | --- |
| Instructor | Tom | Gardos | tgardos@bu.edu |
| Spark Advisor |  |  |  |
| Program Lead | Daniel | Oh | danoh@bu.edu |
| PM | Jiahe (Herbert) | Zhang | hzjh@bu.edu |
| TPM | Julissa | Mijares | jmijares@bu.edu |
| **Primary Client** | Kate | Bergstrom | [kbergs@bu.edu](mailto:kbergs@bu.edu) |
| Teammate | Yuchen | Li | nikkili@bu.edu |
| Teammate | Shiyi | Chen | shiychen@bu.edu |
| Teammate | Gukai | Chen | stevenu@bu.edu |

**Client Notes Section (organized newest to oldest)**

\*Internal Team Notes should be taken in Notion

##### ***1st Client Meeting Template***

Date: 9/25

Attendance: Kate, Herbert, Julissa, Daniel

Relevant Links: *Link to docs relevant to the conversation including most recent client PPT, etc.*

Agenda / Notes:

* Welcome & Introductions - *Use this time to get to know the project management team and client*
* Summary of X-Lab
  + This project is a part of our XLab practicum course [insert course here]. The students working on this project are registered for the course and earning credit for their work and participation. Providing the client with meaningful deliverables is a central output of the class.
* Client Summary of Project
  + In your own words, what is the project? What are you hoping to get out of this semester? Why is this project important for your organization?
* Review of Project Description
  + Confirmation of priorities and deliverables
    - Commenting on PD where adjustments need to be made, if any
    - Goal is to be able to reiterate these priorities to the student team next week
  + Questions from PM / TPM
    - What are some of the directions you want the students to focus on?
* Logistics & Timeline
  + Communication
    - Contact information - is the above contact information correct?
    - What’s the best way to contact you with project questions? Email or would you like to be added to a project slack channel?
    - Any other preference re: communication?
  + Semester Schedule
    - Students are diving into the work now, getting to know the project through the PD and meetings with the PM and TPM
    - Mid October (10/18): about ½ way through the semester
    - Thanksgiving Recess: November 27-December 1
    - Last Day of classes for students: December 10
      * Final presentations to clients will be done before December 10th
    - Spark! Demo Day: December 11th (Clients will be invited to see final posters from teams)
  + Client Meetings
    - We’ll meet for client meetings (with the student team joining after this point) about every other week. Client schedules and holidays may impact this, but every other week is the goal.
    - At client meetings the teams will present their progress updates, ask questions, and ensure alignment to your goals
      * Our goal is to have client meetings towards the end of the week so that the team can meet internally at the start of each week to ensure a high quality presentation for clients.
    - **Looking ahead to the rest of the client meetings, does this day of the week/time of day typically work for you? If not, any preferences for days/times moving forward?** 
      * I will be looping the team into the scheduling of future meetings, so knowing your general availability on a given day/time is helpful.
* Questions / Notes

Objection: correlations/patterns between regions/demographics/origins and students’ academic performance. For example, students recruited from Canada with a high school GPA of 3.5 might not struggle for academics in BU (to get a 3.5 or above), yet students recruited from American high schools only need a 3.2 GPA.

Satisfied with the current timeline.

Meeting time: Wednesdays at 9:30 works, also need to follow up with whether we could meet on Fridays.

##### ***Client Meeting***

Date: OCT 11th

Attendance: Kate, Herbert, Julissa, Daniel, Yuchen, Shiyi, Gukai

Agenda / Notes:

* Welcome
* Project Overview (*what stage of the project is the team in?)*
* Demo
  + Updates since last meeting
    - Preliminary analysis and data clearings are done.
  + Walk through of findings and process
  + What’s next
* Blockers / Questions for client
* Client feedback and questions
* Logistics
  + Confirmation of next meeting
  + Confirmation of action items ahead of next meeting
* NOTES from clients
  + Currently in good process.

##### ***Client Meeting***

Date: OCT 25th

Attendance: Kate, Herbert, Julissa, Daniel, Yuchen, Shiyi, Gukai

Agenda / Notes:

* Project Overview: Data
* Demo
  + Updates since the last meeting
    - Finished the Dataset cleaning and preliminary analysis
  + Walk through of findings and process
  + Future plan, finish half of the question by the next meeting
* Blockers / Questions for client
* Client feedback and questions
* Logistics
  + Confirmation of next meeting
  + Confirmation of action items ahead of next meeting

##### ***Client Meeting***

Date: NOV 15th

Attendance: Kate, Herbert, Julissa, Daniel, Yuchen, Shiyi, Gukai

Agenda / Notes:

* Update: Students has finished all tests
* Blockers / Questions for client
* Client feedback and questions
* Logistics
  + Confirmation of next meeting
  + Confirmation of action items ahead of next meeting

##### ***Client Meeting***

Date: NOV 22th

Attendance: Kate, Herbert, Julissa, Daniel, Yuchen, Shiyi, Gukai

Agenda / Notes:

* Welcome
* Project Overview:
* Demo
  + Updates since last meeting
    - Finished the
  + Walk through of findings and process
  + What’s next
* Blockers / Questions for client
* Client feedback and questions
* Logistics
  + Confirmation of next meeting
  + Confirmation of action items ahead of next meeting

##### ***Client Meeting***

Date: Dec 10th

Attendance :Kate, Herbert, Julissa, Daniel, Yuchen, Shiyi, Gukai

Agenda / Notes:

* Welcome
* Project Overview:
* Demo
  + Updates since last meeting
    - Finished the
  + Walk through of findings and process
  + What’s next
* Blockers / Questions for client
* Client feedback and questions
* Logistics
  + Confirmation of next meeting
  + Confirmation of action items ahead of next meeting

### **Semester Wrap-Up Overview**

1. Tasks Complete this Semester
   1. Students cleaned datasets tailored to their specific analytical needs.
   2. Conducted single-variable statistical analysis and multivariable statistical analysis.
   3. Explored correlations between geographic regions and students' GPA academic performance.
   4. Students analyzed basic statistical information about student-athletes' data
2. Recommended Next Steps
   1. Gather additional data from the client to enhance the analysis.
   2. Perform verification tests on current results to ensure statistical significance.
   3. Expand future analysis with more in-depth multivariable techniques.
3. Comments/Insights (Best Practices: Tips for next person to be successful)
   1. If there are things in mind that you want to assign the students to do, then be specific and provide sufficient details so that the students don’t get lost.
   2. It is more effective to assign each student something rather than letting them figure out what they should do.
   3. Let the students know that the clients’ preference is more important than what it says in the documentation.