

Project Plan: Potential Season Ticket Holder Algorithm

By: AJ Johnson

Objective

To develop an algorithm that identifies potential season ticket holders or dedicated fans using various factors like alumni status, social media activity, game attendance history, and merchandise purchases.

Tools & Resources Needed

1. Database Solution
 - a. To store collected data. (SQL or NoSQL)
2. Data Collection Tools
 - a. Web scrapers, API's for social media, and integration with existing systems in the universities ocean of data.
3. Machine Learning Library
 - a. Libraries and packages used to train models of this nature
4. Computing Resources
 - a. A decent server or cloud resources to run the machine learning algorithms.
5. Dashboard Software (Preferred)
 - a. Used to display query results in an easy-to-understand way.

Initial Risk Assessment

1. Data Privacy
2. Cost
 - a. Computing resources and API costs.
3. Algorithm Accuracy
 - a. Its only as good as the data its trained on.

Milestones

1. Feasibility Study & Approval
2. Hardware & Software Planning
3. Development
4. Installation
5. Pilot Testing
6. Launch
7. Review & Iterate

Hardware

Depending on the needs, the hardware requirements can vary. If the service needs to be constantly running and reassessing the algorithm, then the hardware should be reliable and relatively up-to-date to ensure performance.

On the other hand, if the application is designed to be run on-demand—say, when a marketing team wants to identify potential customers for merchandise—they could simply go to the application, click "Run," and wait for the results. In this scenario, the hardware doesn't need to be as sophisticated, as it's not expected to be running continuously or constantly reassessing the data.

Final Thoughts and Concerns

My concerns about this project are numerous, but they should be manageable. The top issues I foresee relate to model validation (accuracy), inter-departmental collaboration (data acquisition), and scalability.

I can't fully address model validation at this stage because I'm uncertain about the data landscape. Specifically, I'm unsure how much data is accessible or stored, and how much of it would be usable for the model.

This leads me to my next point: data sensitivity and diversity. Data across different departments—like merchandise and ticket sales—are likely stored differently due to their unique needs. Harmonizing this data into a format that's ingestible by an algorithm will require extra development time. It's not an insurmountable challenge, but it will take time. After all, a model is only as good as its data.

Finally, I have concerns about scalability. While this may not be an immediate issue, its impact will be significant down the line. Scalability will influence whether the application can run on a local machine or will require a dedicated cloud server. While this might seem exaggerated, effective scaling could make this tool invaluable for multiple staff members. It could even be used to generate charts and data for internal meetings where such insights are needed.