Sports Analytics Bot Final Report

Team CS Enthusiasts

Karan Reddy Kandala, Soham Patil, Aidan Walters, Aneesh Tummeti
Department of Computer Science
Virginia Tech
Blacksburg, VA
Karanreddyk@vt.edu, sohampatil@vt.edu, aidanw@vt.edu, artummeti@vt.edu

Abstract

In sports analytics, the increasing complexity of data needs innovative bots capable of analyzing large datasets to extract meaningful insights for sports coaches, players, and personnel. Our proposed bot leverages state-of-the-art machine learning and natural language processing to automate data analysis, such as player metrics. This tool aims to improve strategic decision-making, enhance player performance, and boost fan engagement by utilizing the power of data analysis in Python, tools like BeautifulSoup, Flask, and SQL databases, in conjunction with real-time sports data from APIs such as ESPN's. Using Agile methodologies, our project progressed through cycles of planning, execution, and evaluation to meet objectives within this semester.

Introduction

The need for a capable bot to handle data analytics is growing in the fast-paced field of sports analytics. Sports organizations encounter difficulties in analyzing and obtaining practical insights due to the abundance of data sources, which span player performance statistics. A specialized bot with cutting-edge machine learning and natural language capabilities could automate this process, simplify data management, and give coaches and staff real-time insights. A bot like this may provide sports with a competitive edge by using data analytics to improve fan experiences, optimize athlete performance, and enable data-driven decisions for all. Sports betting is also one of the fastest growing industries [4].

Problem & Motivating Example

The problem we are looking to solve is the lack of easy access to sports data and the difficulty of making decisions related to sports. An example of someone using this bot is a General Manager in the NFL who is planning who they are going to pick in the NFL draft. They might watch a film to analyze a player's technique, but they will also need to see statistics pertaining to a certain position. If they were drafting a receiver, they could see stats such as players who had the most receiving yards, minimized drops, and the highest rate of being open. They could search up these stats and then the bot could show them which players excelled in each of these categories and any other statistics.

Background

1. Streamlit: Streamlit is an interactive web applications tool for python. Its ease of use enabled us to make quick prototypes for the product. We were able to create

intuitive interfaces that allowed us to make a product that can interact without major issues.

2. Large Language Model (LLM): A large language model (LLM) is a type of artificial intelligence system designed to understand, generate, and manipulate human language based on patterns learned from vast amounts of textual data. These models use deep learning techniques, particularly transformer architectures, to process and produce text in a way that is contextually relevant and often indistinguishable from human-written content.
3. OpenAI API: offers natural language processing models, including large language models (LLM's). We integrated this into the project to allow us to analyze text-based responses from the user and produce accurate results from them.
4. Llama 2 LLM (Large Language Model): Llama 2 LLM by Meta is a very specific model that processes language to produce data. It can process information at a high rate, especially in the context of sports analytics for this project.
5. Kaggle: Kaggle offers a wide selection of datasets to train AI models. To create a bot for data analytics in sports, we used Kaggle datasets to gather information and train the bot.
6. VSCode: VSCode is a very popular IDE (integrated development Environment) and works well with various programming languages.

Related Work

Notably, platforms such as IBM's Watson Analytics and SAP Sports One have set industry benchmarks by employing advanced analytics to deliver predictive insights into player performance and game strategies. These systems integrate diverse data sources and apply complex statistical models to enhance team management and performance evaluations. Despite their robust capabilities, many of these platforms do not fully address the need for real-time data analysis during live sports events, which is critical for immediate strategic decision-making. Our project introduces a novel approach by integrating real-time data processing with interactive features, leveraging the Python ecosystem with tools like BeautifulSoup for efficient web scraping and Flask for seamless web application functionality. This innovation is further enhanced by the utilization of the OpenAI API and the Llama 2 Large Language Model for natural language processing, enabling our bot to interpret and analyze text-based data dynamically. This capability allows for instant analysis and response during live games, providing a competitive edge not readily found in current market offerings. Moreover, the practical application of our solution in real-world scenarios is underscored by its ability to interface directly with popular sports data APIs such as ESPN's. This integration enables the bot to pull live data feeds, ensuring that the analysis reflects current game conditions and player performances.

Implementation Plan

Every software project requires an effective process to accomplish tasks efficiently and properly. For our project, we utilized the Agile Approach [1]. Given that we had less than a semester to complete this project, we aimed to maximize our productivity. The agile process provided us with the necessary flexibility, which was crucial as we were initially uncertain about the exact implementation details of the project. This flexibility

allowed us to adapt and make changes as needed. We organized our project work as follows:

- We started with planning, which involved sprints or meetings where we discussed our immediate tasks and objectives.
- For the design phase, our team explored how we wanted the backend and frontend of the project to look.
- Next, we moved into development, where we put our plans into action and began coding.
- Testing was an essential phase where we ensured that the bot produced the expected results.

Finally, we conducted a review of our work and used the insights gained to plan subsequent phases of the project. Utilizing the Agile process helped us complete project components efficiently. This structure also facilitated breaking down the problem into manageable parts, simplifying the overall execution. Initially, we established a GitHub repository, providing all group members with a platform to plan and access the project collaboratively. GitHub enabled us to create tasks, track bugs or issues, and fostered a collaborative environment for project implementation.

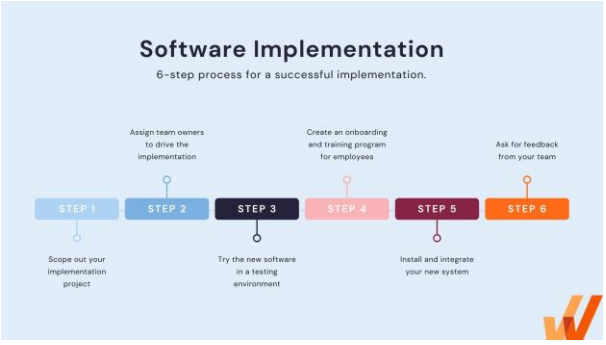

Image 1: Software Implementation Process

After setting up our GitHub repository, the next step was to select an Integrated Development Environment (IDE) for coding. We opted for VS Code because of its robust integration with GitHub and the ease of downloading necessary dependencies. The familiarity of our group with VS Code also influenced this decision, making it a practical choice for our project. We planned to integrate Large Language Models (LLMs) into our project to facilitate data analysis. Given the complexities of handling sports data through sequential programming, LLMs like OpenAI and Llama2 provided a dynamic framework that greatly enhanced our analytical capabilities.

Once satisfied with the backend setup, our focus shifted to the front-end implementation. We conducted extensive research and examined existing sports analytics applications to inform our design. Ultimately, we chose a straightforward approach using HTML and CSS, which not only required minimal resources but also seamlessly connected with our bot. In terms of design, we adopted a data-centric architecture to ensure continuous interaction with Llama2 throughout the project. Additionally, testing was crucial to verify the accuracy of our bot's responses. Testing AI-driven components poses unique challenges due to their potential unpredictability, necessitating thorough validation to ensure reliability.

Deployment & Maintenance Plan

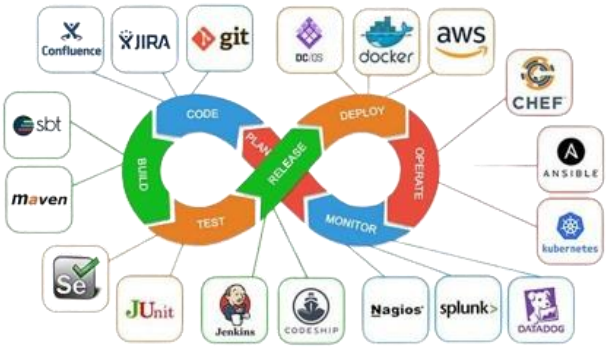

Image 2: Deployment Tools (DevOps)

In the deployment of our sports analytics bot, Streamlit Cloud emerges as a prime candidate for hosting our application. Leveraging Streamlit Cloud has distinct advantages; it is tailor-made for Streamlit apps (the app we used to host our demo), offering seamless integration from development to deployment using GitHub, the code management platform we used. Its ease of use is one of its strongest points, allowing developers to swiftly transition from a prototype to a live application accessible by users worldwide. This platform automates deployment processes, thus adhering to the Agile methodology of continuous delivery and iterative feedback. However, one must consider potential limitations, such as dependency on external hosting services, which could introduce concerns about uptime guarantees and the handling of high-traffic scenarios.

Beyond initial deployment, we'll establish robust monitoring and logging practices to oversee the application's performance. Utilizing Docker for containerization guarantees that the bot runs consistently across all deployments, and a CI/CD pipeline will ensure a smooth transition from repository to production. Maintenance will be active, focusing on regular updates, scalability to manage growing datasets of sports analytics and user bases, and strict security measures to protect sensitive data. Such a structured approach will facilitate rapid, agile responses to any technical or user experience issues that arise in the future.

The final aspect of our deployment and maintenance strategy includes a detailed user support and a go-to-market plan. We plan on having clear documentation and user training as this will be pivotal for adoption, providing clear guidance on leveraging the bot's features. Feedback channels will enable iterative enhancements, ensuring the bot evolves to meet user demands. If our vision extends to commercializing our project, we'll devise a strategy to engage sports organizations like ESPN and CBS, possibly exploring partnerships to increase our reach and impact. With Streamlit Cloud as the starting block in our strategy, our deployment plan is designed to be scalable, secure, and user-centric, paving the way for a sports analytics tool that is as dynamic as the data it analyzes.

Discussion & Future Work

If future work were to be conducted on this project, we could incorporate more advanced machine learning techniques. Currently, our implementation primarily utilizes a Large Language Model (LLM), but we could expand this to include a bespoke model specifically tailored for sports analysis. Such a model, designed for the nuances of sports data, could significantly enhance the accuracy of our predictions and broaden the range of functionalities our bot can handle. Additionally, we could enhance the user interface (UI) of our project. At present, our front end is relatively simple and lacks extensive functionality. Future enhancements could involve adding more interactive sections to

the website, such as leaderboards for top scorers and additional statistics that users might find valuable.

During our discussion presentation, we enjoyed showcasing our project to the class and receiving feedback. This session allowed us to engage with the audience, respond to their inquiries, and demonstrate the progress we achieved over the semester. The insights gained from this interaction are invaluable for the continuous improvement of our project, suggesting areas where additional features and refinements are needed. This ongoing dialogue with potential users underscores the dynamic nature of development in sports analytics, highlighting both the challenges and opportunities for technological innovation in this field.

Conclusion

In conclusion, our project successfully developed a sports analytics bot designed to enhance data-driven decision-making within the sports industry. By leveraging state-of-the-art machine learning models and natural language processing techniques, our bot effectively automates the analysis of comprehensive sports data. However, despite achieving our primary objectives, our project faced several limitations. Time constraints were a significant challenge, as the semester provided a limited window to develop and refine our system fully. Additionally, resource availability restricted the extent of features and functionalities we could implement within this timeframe.

Looking ahead, future enhancements could include the integration of more advanced machine learning techniques to refine the bot's predictive accuracy and expand its analytical capabilities. Improving the user interface would also be a priority, potentially incorporating a more interactive and visually engaging platform that offers real-time updates and more detailed statistical breakdowns. Such developments would ensure that our sports analytics bot not only meets but exceeds the evolving needs of the sports industry, thereby cementing its value in enhancing competitive strategies and engaging sports audiences more effectively.

References

[1] S. Laoyan, "What is agile methodology? (a beginner's guide) [2023] • asana," Asana, https://asana.com/resources/agile-methodology (accessed Feb. 18, 2024).

[2] American University, "How Analytics Is Changing Sports [2023] • American University," American University, https://programs.online.american.edu/mssam/sports-management-masters/resources/how-analytics-is-changing-sports (accessed April 2, 2024).

[3] CloudFactory, "Sports Analytics with High-Quality Data Labeling [2023] • CloudFactory Blog," CloudFactory Blog, https://blog.cloudfactory.com/sports-analytics-with-high-quality-data-labeling#:~:text=Machine%20learning%20(ML)%20applications%20in,fans%2C%20and%20everything%20in%20between. (accessed April 23, 2024).

[4] M. Sexton, "USC researcher studies the growth of Sports Gambling," University of South Carolina, https://sc.edu/uofsc/posts/2024/01/stephen-shapiro-sports-gambling.php (accessed May 2, 2024).