

فاز اول پروژه: استخراج داده‌های بسکتبال از Basketball Reference

در فاز اول این پروژه، تمرکز اصلی بر جمع‌آوری داده‌های خام از وبسایت معتبر **Reference Basketball** بوده است. این داده‌ها شامل اطلاعات بازیکنان، تیم‌ها، قهرمانان فصل‌ها و لیست بازیکنان جایزه **MVP** بوده و هدف از این مرحله، آماده‌سازی یک مجموعه داده جامع و قابل اعتماد برای مراحل بعدی طراحی پایگاه داده و تحلیل‌های آماری بوده است.

برای انجام این کار، از ترکیبی از کتابخانه‌های قدرتمند پایتون استفاده شده است. کتابخانه‌های **requests** و **BeautifulSoup** برای دریافت و پارس کردن صفحات **HTML** به کار گرفته شدند، در حالی که **selenium** برای استخراج اطلاعات از صفحات داینامیک و تعامل با DOM واقعی مورد استفاده قرار گرفت. همچنین **pandas** برای ساخت دیتافریم‌ها و ذخیره‌سازی داده‌ها به صورت فایل‌های **CSV** استفاده شد. برای پردازش‌های دقیق‌تر مانند استخراج قد، وزن، ملیت، تجربه و تاریخ تولد یا مرگ بازیکنان، از **re** و **datetime** بهره گرفته شد.

در این فاز، ابتدا لیست ۶۰ بازیکن برتر هر فصل از فصل **20-2019** تا **25-2024** استخراج شد. سپس بازیکنان تیم مربوط به قهرمان فصل‌های **20-2019** تا **25-2024** جمع‌آوری گردید. در ادامه، با استفاده از **Selenium**، اطلاعات بیوگرافی و آماری هر بازیکن از جمله سن، قد، وزن، ملیت، تیم فعلی، پوزیشن بازی، تجربه حرفه‌ای و میانگین امتیاز (**PTS**) استخراج شد. همچنین اطلاعات مربوط به لیست بازیکنان جایزه **Michael Jordan Trophy** برای فصل‌های **20-2019** تا **25-2024** جمع‌آوری گردید. در پایان، اطلاعات پایه‌ای تیم‌های فعال لیگ شامل نام تیم، موقعیت جغرافیایی، سال تأسیس، تعداد قهرمانی‌ها و تعداد حضور در پلی‌آف نیز استخراج شد.

تمام داده‌های جمع‌آوری‌شده در قالب فایل‌های **CSV** ذخیره شده‌اند تا در فاز دوم پروژه، یعنی طراحی پایگاه داده، مورد استفاده قرار گیرند.