



০০১-০০৩ ডেটা সায়েন্সে পাইথনের ভূমিকা

Statistics কে কিছুটা বিরতি রেখে আমরা এখন পাইথন ধরছি, পরে আবার আমরা *Statistics continue* করবো।

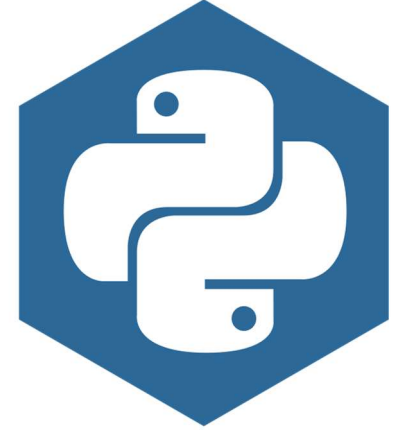
ডেটা সায়েন্স হলো একটি বহুমুখী ক্ষেত্র যেখানে বড় পরিমাণের তথ্য সংগ্রহ, বিশ্লেষণ ও মডেলিং করে সিদ্ধান্ত গ্রহণ করা হয়। এই কাজে পাইথন (Python®) অত্যন্ত জনপ্রিয় একটি প্রোগ্রামিং ভাষা, যা সহজ সিনট্যাক্স ও শক্তিশালী লাইব্রেরির জন্য ডেটা সায়েন্টিস্টদের পছন্দের প্রথম সারির টুল।

Attribution Statement (for PSF)

Python® is a registered trademark of the [Python Software Foundation \(PSF\)](#). This document is an independent work and is not affiliated with, endorsed by, or sponsored by the [Python Software Foundation \(PSF\)](#). This logo is a **custom adaptation** of the Python logo, designed for **illustrative** and **educational use**.

Got Python snake SVG from:

<https://www.svgrepo.com/svg/512738/python-127> (PD License)



ডেটা সায়েন্সে পাইথনের প্রধান ভূমিকা

১. ডেটা সংগ্রহ ও প্রক্রিয়াকরণ: ডেটা সায়েন্সের প্রথম ধাপ হলো বিভিন্ন উৎস থেকে ডেটা সংগ্রহ করা এবং সেটিকে বিশ্লেষণের উপযোগী করে গঠন করা। পাইথনের নিম্নলিখিত লাইব্রেরিগুলো এতে সাহায্য করে:

Pandas – টেবুলার ডেটা পরিচালনার জন্য ব্যবহার করা হয়

NumPy – গাণিতিক ও সংখ্যাসূচক বিশ্লেষণের জন্য

BeautifulSoup, Scrapy – ওয়েব স্ক্র্যাপিং করে ডেটা সংগ্রহের জন্য

২. ডেটা বিশ্লেষণ ও ভিজুয়ালাইজেশন: ডেটা বিশ্লেষণ করে তা থেকে গুরুত্বপূর্ণ তথ্য বের করতে পাইথনের বিভিন্ন টুল ব্যবহৃত হয়:

Matplotlib – গ্রাফ ও চার্ট তৈরির জন্য

Seaborn – উন্নতমানের ডেটা ভিজুয়ালাইজেশনের জন্য

Plotly – ইন্টারঅ্যাকটিভ চার্ট তৈরির জন্য

৩. মেশিন লার্নিং ও কৃত্রিম বুদ্ধিমত্তা: ডেটা সায়েন্সের অন্যতম গুরুত্বপূর্ণ অংশ হলো মেশিন লার্নিং, যা পাইথনের মাধ্যমে সহজেই করা যায়:

Scikit-learn – মেশিন লার্নিং অ্যালগরিদম ব্যবহার করতে

TensorFlow, PyTorch – ডিপ লার্নিং ও নিউরাল নেটওয়ার্ক তৈরির জন্য

৪. বড় ডেটার বিশ্লেষণ (Big Data Processing): বড় আকারের ডেটা বিশ্লেষণের জন্য পাইথন ব্যবহৃত হয়:

Dask – বড় ডেটাসেট প্রসেস করতে

PySpark – Apache Spark এর মাধ্যমে ডিস্ট্রিবিউটেড ডেটা প্রসেসিং

৫. স্বয়ংক্রিয়করণ ও ডিপ্লয়মেন্ট: পাইথন ডেটা সায়েন্স মডেল স্বয়ংক্রিয় করতে এবং প্রোডাকশনে ব্যবহৃত হয়:

Flask, FastAPI – মডেল ডিপ্লয়মেন্টের জন্য

Airflow – ডেটা পাইপলাইন অটোমেশন করতে

উপসংহার

পাইথন তার সহজলভ্যতা, বিশাল লাইব্রেরি সংগ্রহ, ওপেন সোর্স কমিউনিটি এবং শক্তিশালী বিশ্লেষণ ক্ষমতার জন্য ডেটা সায়েন্সের অপরিহার্য হাতিয়ার হয়ে উঠেছে। এটি ডেটা সংগ্রহ থেকে শুরু করে মডেল ডেভেলপমেন্ট ও ডিপ্লয়মেন্ট পর্যন্ত সমগ্র প্রক্রিয়াটি সহজ করে তোলে।

অতএব, পাইথন শেখা মানেই ডেটা সায়েন্সে এক ধাপ এগিয়ে যাওয়া!