

ইলিয়াম এডওয়ার্ড ডেমিং বলেছেন- ‘In God we trust; all others must bring data’। তার এই কথার মর্মার্থ হচ্ছে- ঈশ্বরকে মানার বিষয়টি বিশ্বাসনির্ভর; আর বাকি সবকিছুতেই প্রয়োজন ডাটা। এ লেখায় আমরা উদঘাটন করার প্রয়াস পাব, কী করে কোনো সিদ্ধান্ত গ্রহণে বৈজ্ঞানিকভাবে আমরা ডাটা ‘bring’ করি, তথা ডাটাকে নিয়ে আসতে পারি।

এটি দেখতে খুবই অস্বাভাবিক লাগে, এ ক্ষেত্রের অনেক প্র্যাকটিশনার সবচেয়ে হালনাগাদ ও বড় বড় টুল ব্যবহার করেন অনেক বড় ও জটিল ডাটাসেটে। এরা দেখতে পান তাদের ফলাফল সিদ্ধান্ত-নির্ধারণের তথা ডিসিশন মেকারেরা বাতিল করে দেন। কারণ, এখানে ডোমেইন সায়েন্সের সমস্যার সমাধান করা হয়নি। অনেক বড় বড় সিদ্ধান্ত নেয়া হয় সঙ্কীর্ণ ভাবনা-চিন্তা থেকে, সেখানে থাকে না ডাটাভিত্তিক প্রক্রিয়া (ডাটা ড্রিভেন প্রসেস)। এমনকি

হয় একটি সিস্টেম্যাটিক এন্টারপ্রাইজ, যা সৃষ্টি করে, গড়ে তোলে ও সংঘটিত করে পরীক্ষণযোগ্য জ্ঞান। আর জগৎ সম্পর্কে ভবিষ্যদ্বাণী পদক্ষেপগুলোতে অবশ্যই অপরিহার্যভাবে অন্তর্ভুক্ত থাকতে হবে তিনটি বিষয়-

০১. বিশ্লেষণধর্মী বিশ্লেষণ : বিশ্লেষণধর্মী বিশ্লেষণ বলতে আমরা বুঝব অ্যানালাইটিকের মাধ্যমে অ্যানালাইসিসকে। আমরা একটি সমস্যা সমাধানকল্পে সংজ্ঞায়িত করি একটি সুনির্দিষ্ট সমস্যা, প্রাথমিক ডাটা সংগ্রহ করি বিভিন্ন সেসিং মেকানিজম ব্যবহার করে, পূর্বানুমান বা হাইপোথেসিস তৈরি করি এই সমস্যাসংশ্লিষ্ট দুনিয়াটা কীভাবে কাজ করে, হাইপোথেসিস পরীক্ষা-নিরীক্ষার জন্য এক্সপেরিমেন্ট ডিজাইন করি, পদ্ধতিগতভাবে ডাটা সংগ্রহ করি, পরিসংখ্যানগত ও মেশিন-লার্নিং কৌশল ব্যবহার করে আমাদের হাইপোথেসিস পরীক্ষা করি এবং

মডেলের মাধ্যমে, জানতে পারি কী পদক্ষেপ নিতে হবে (হোয়াট-টু-ডু) অপটিমাইজেশনের (হোয়াট ইজ বেটার) মাধ্যমে।

ডাটা সায়েন্সের অর্থ সবার আগে বিজ্ঞান

আমরা দেখেছি- এই বৈজ্ঞানিক প্রক্রিয়ার প্রেসক্রিপশন অনুসরণ করে বড় ধরনের অগ্রগতি এনে দিয়েছে ইলেকট্রনিকস, জেনোমিকস, কেমিস্ট্রি, মেকানিকস, এয়ারোনটিকস ও বিজ্ঞানের অন্যান্য বিষয়ের টেকনোলজিতে। এই প্রক্রিয়া প্রয়োগ করা হয়েছে ভৌতবিজ্ঞান ও জীববিজ্ঞানের নানা ক্ষেত্রে। এখন আমরা বিজ্ঞানের এই একই প্রক্রিয়া ব্যবহার করছি ব্যাংকিং সায়েন্স, ইনভেস্টমেন্ট ম্যানেজমেন্ট সায়েন্স, হিউম্যান ক্যাপিটাল ম্যানেজমেন্ট সায়েন্স, কাস্টমার রিলেশন-শিপ ম্যানেজমেন্ট সায়েন্স, সাপ্লাই চেইন ম্যানেজমেন্ট সায়েন্স, ম্যানুফেকচারিং ম্যানেজমেন্ট সায়েন্স, অ্যাসেট ম্যানেজমেন্ট সায়েন্স, ফিন্যান্সিয়াল ফ্রন্ট সায়েন্স, অপারেশন রিসার্চ, অর্গানাইজেশনাল অ্যান্ড বিহেভিয়ারেল সাইকোলজি, গেম থিওরি ইত্যাদি নতুন নতুন ক্ষেত্র সৃষ্টি করতে। একটি বৃহত্তর ডোমেইন সেটে বৈজ্ঞানিক পদ্ধতির এই প্রয়োগ এখন একটি ইন্ডাস্ট্রি শর্টহ্যান্ড হিসেবে পরিচিত ‘ডাটা সায়েন্স’ নামে। এসব প্রতিটি ডোমেইনের বিজ্ঞানীদের ভালো করে জানতে হবে তাদের ডোমেইন সম্পর্কে, বুঝতে হবে ডাটাসেটের বিভিন্নতা সম্পর্কে এবং এই জ্ঞানকে প্রয়োগ করতে হবে তাদের ডোমেইনের আদর্শ প্রেসক্রিপশন বের করে আনার জন্য।

এখানে উপস্থাপিত হলো এই বিজ্ঞান প্রক্রিয়ার ব্লুপ্রিন্টের জন্য একটি প্রস্তাব, যা প্রয়োগ করা হয় মাল্টিপল ডোমেইনে এবং যা কাজ করে উল্লেখযোগ্য সাফল্যের সাথে।

০১. প্রথমেই ডিজাইন থিঙ্কিংয়ের মাধ্যমে সংজ্ঞায়িত করুন মুখ্য বিজনেস প্রবলেম।

আমাদের অভ্যন্তরীণ/বাহ্যিক গ্রাহকের প্রত্যাশা কী?

ডিফারেন্সিয়েশনটা কী? আমরা কি বিক্রি করতে পারি? কোনটি মূল্যবান?

প্রযুক্তির দিক থেকে কোনটি বাস্তবায়নযোগ্য? বিদ্যমান অভ্যন্তরীণ ও বাহ্যিক প্রযুক্তির ব্যবহার।

বিজনেস হিসেবে কোনটি আমাদের জন্য টেকসই?

সময়/অর্থ ও সমাধানের জটিলতার মধ্যে ভারসাম্য আনা।

উচ্চমান ও নির্ভরযোগ্যতা বনাম পিওসি পর্যায়ের পদক্ষেপ।

উদঘাটনে বিনিয়োগ ও আইপি সৃষ্টিতে উপায় বের করা।

আমরা কী উন্নয়ন করতে পারি ও তাতে কি করে গতি আনতে পারি?

০২. ডাটা অ্যানালাইটিক হচ্ছে টিমওয়ার্ক। আর এই টিমওয়ারকে প্রয়োজন কমপক্ষে তিনটি ভূমিকা। এজন্য দরকার বিশেষায়িত প্রশিক্ষণ।

সিকিউরিটি ও প্রাইভেসির জন্য ডাটা স্টুয়ার্ড, ▶

ডাটা সায়েন্সে বিজ্ঞানটা কোথায়?

গোলাপ মুনীর

যখন সিদ্ধান্ত-প্রণেতা ডাটা অনুসরণ করেন, তখন তাদের থাকে এমন তত্ত্ব বা পূর্বানুমান, যা এরা যাচাই করেন সঙ্কীর্ণ উপলব্ধি নিয়ে এবং সে অনুযায়ী বাজেট বরাদ্দ করেন। অতএব ব্যবস্থাপকেরাই কোনো কোনো সময় সংজ্ঞায়িত করেন- কোন পরীক্ষা করা হলো, কোন ডাটা সংগৃহীত হলো কোনো ডাটাভিত্তিক প্রক্রিয়া ছাড়াই।

কেনম হতো, যদি আমরা কাজ করতাম বিজনেস সমস্যাকে সামনে রেখে, শুরুটা করতাম ডোমেইন নলেজ নিয়ে, পাওয়া ডাটার ওপর ভিত্তি করে সংজ্ঞায়িত করতাম পরীক্ষা ও পূর্বানুমান তথা হাইপোথেসিস, আর এভাবে সিদ্ধান্ত-প্রণেতাদের সহায়তা দিতাম ডাটাভিত্তিক সিদ্ধান্ত গ্রহণে?

বিজ্ঞানের প্রক্রিয়া

পৃথিবীটা সম্পর্কে যথাসম্ভব বেশি জানা ও এই জানা বা জ্ঞানকে নিজের স্বার্থে কাজে লাগানো মানবজাতির দীর্ঘদিনের এক প্রত্যাশা। কখনও কখনও ব্যবহার করা হয় সংখ্যাগত পদক্ষেপ। আজ আমাদের ডাটা কালেকশন সিস্টেম সম্পর্কে মনে হয় আমরা অনেক কিছুই জানি, কিন্তু এরপরও এর মূল্য সম্পর্কে জানি খুবই কম। আমরা যদি জানি- আমাদের দুনিয়াটা কীভাবে চলে বা কাজ করে, তবে আমরা এর সম্পর্কে যুক্তিসঙ্গত ভবিষ্যদ্বাণী করতে পারব, একে কার্যকরভাবে ব্যবহার করতে পারব অর্থনৈতিক উপকার বয়ে আনার ক্ষেত্রে। এই জগতটা হতে পারে গোটা মহাবিশ্ব, অথবা এটি হতে পারে আমাদের ছোট্ট এন্টারপ্রাইজ বিজনেস। পৃথিবীটা কীভাবে কাজ করে, তা আবিষ্কার করার একটি কৌশল হতে পারে এমপিরিক্যাল সায়েন্স তথা বাস্তব পর্যবেক্ষণ ও পরীক্ষা-নিরীক্ষানির্ভর বিজ্ঞান। বিজ্ঞানে প্রয়োজন

মডেল ও পরীক্ষণের মাধ্যমে আমাদের হাইপোথেসিস সংশোধন-পরিশোধন করি নতুন একটি হাইপোথেসিস পাওয়ার জন্য।

এখানে ডাটা বিশ্লেষণের উপাদানে প্রয়োজন হয় বিভিন্ন ধরনের ডাটা ইমপোর্টিং, সংশ্লিষ্ট অংশটি বের করে আনা, ট্রান্সফর্মিং ও লোড। এরপর আমাদের প্রয়োজন সমস্যাসংশ্লিষ্ট ডাটা থেকে বৈশিষ্ট্য চিহ্নিত ও সৃষ্টি (আইডেন্টিফাই ও ক্রিয়েট) করা, ট্রেনিং ও টেস্টসেটের জন্য একটি স্যাম্পলিং স্ট্র্যাটেজি সংজ্ঞায়িত করা, সুনির্দিষ্ট মেশিন-লার্নিং বা পরিসংখ্যানিক সূত্রায়ন, মডেল প্যারামিটার সেলেক্টিভিটির জন্য মন্টি-কার্লো অপটিমাইজেশন চালু করা এবং মডেলগুলোর ক্রস-ভ্যালিডেটিং করা।

০২. মডেলিংয়ের মাধ্যমে সংশ্লেষণ : মডেলিংয়ের মাধ্যমে সংশ্লেষণের (সিনথেসিস থ্রো মডেলিংয়ের) বেলায় আমরা জগত-সম্পর্কিত একটি তত্ত্ব তৈরি করতে জ্ঞানকে ব্যবহার করি আউট টেস্টেড হাইপোথেসিস থেকে, যাতে অন্তর্ভুক্ত করা যেতে পারে মেশিন-লার্নিং/ভিত্তিক মডেল। হতে পারে আমাদের মডেলগুলোকে ক্রস-ভ্যালিডেট করতে, মডেলের প্যারামিটার সেলেক্টিভিটি নির্ণয় করতে ও নতুন ডাটাসহ ভবিষ্যদ্বাণী করার জন্য মডেল চালু করতে সহায়তা নেয়া যেতে পারে একটি এক্সপার্ট সিস্টেম, ডায়নামিক গাণিতিক মডেল (পার্শিয়াল ডিফারেনশিয়াল ইকুয়েশন) ও পরিসংখ্যানিক কৌশলের।

০৩. হোয়াট-ইফ সিমুলেশন ও অপটিমাইজেশনের মাধ্যমে প্রেসক্রিপশন : আমরা মন্টি-কার্লো সিমুলেশনের বাইরে ভবিষ্যদ্বাণী তৈরি করতে পারি মডেলের সিমুলেশন (হোয়াট ইফ)

এক্সট্রাক্ট-ট্রান্সফর্ম-লোডিং ডাটা এবং কার্যকর ডাটা ট্রান্সফরমেশন। অবশ্যই ডাটা ব্যবস্থাপনা করতে হবে অ্যালগরিদমিক্যালি রিপ্রেজেন্টেবল প্রোভেন্স (শুধু ডকুমেন্টেড অথবা মেটাডাটা অথবা লিঙ্কড প্রোভেন্স হলে চলবে না)।

ডাটা প্রকৌশলীরা ডাটা পাইপলাইন ক্রিয়েট ও স্কেল করার জন্য আলোকপাত করেন টুল ও আর্কিটেকচারের ওপর।

ডাটা বিজ্ঞানীরা হচ্ছেন ডোমেইনের বিজ্ঞানী, যারা আলোকপাত করেন বিজ্ঞানের ওপর, অর্থাৎ সমস্যা সমাধানের ওপর। এরা জানেন কীভাবে টুল ও আর্কিটেকচার বাড়তি চাপ মোকাবেলা করে। জানেন এর সীমাবদ্ধতাগুলোও। এরা সৃষ্টি করেন ডোমেইনভিত্তিক হাইপোথেসিস ও ভিজুয়লাইজেশন এবং অ্যানালাইটিকগুলো প্রচার করেন ডিসিশন-মেকারদের কাছে।

০৩. ডাটা ইন্টিগ্রেশন ও ক্লিনিংয়েই যায় বেশিরভাগ সময়।

সাধারণত ডাটা পাওয়া যায় মাল্টিপল সোর্সে, যার রিম্যাপ করা প্রয়োজন হয়।

ডাটা থাকে বিভিন্ন মানে, ডাটায় থাকে অসামঞ্জস্যতা, ভ্রান্তি এবং অর্থহীনতা।

ডাটা বিজ্ঞানী, প্রকৌশলী ও স্টুয়ার্ডদের দিয়ে একযোগে কাজ করিয়ে ডাটা রিকনসিল বা অসামঞ্জস্যতা দূর করতে হয়।

কোনো কোনো সময় সমস্যার জন্য গুরুত্বপূর্ণ ডাটা অন্তর্ভুক্ত করা হয় না।

ভবিষ্যদ্বাণী করা শক্ত

সাধারণত ভবিষ্যদ্বাণী করা শক্ত। উঁচুমানের মডেল তৈরি করাও কঠিন কাজ। একটি নলেজ-ড্রিভেন হাইব্রিড (এক্সপার্ট সিস্টেম, ডিফারেনশিয়াল ইকুয়েশনস) এবং ডাটা-ড্রিভেন (স্ট্যাটিস্টিক্যাল, মেশিন-লার্নিং) মডেল দরকার হয় সত্যিকারের প্রিডিকটিভ মডেল সৃষ্টির জন্য। বিশুদ্ধভাবে ডাটা-ড্রিভেন মডেলগুলো বর্তমানের প্রেক্ষাপটে দাঁড়িয়ে শুধু আলোকপাত করে অতীত ডাটার একটি সার-সংক্ষেপ। এরা বুঝতে ব্যর্থ হয় কোথায় পদার্থবিদ্যার বোঝাপড়া এই সমস্যারটির সমাধানে প্রয়োগ করা যেত।

ভিজুয়লাইজেশন ব্যবহার

এখানে প্রয়োজন ভিজুয়লাইজেশন বা দূরদৃষ্টি দিয়ে চেতনা সৃষ্টি করা। আপনি যদি আপনার ভিজুয়লাইজেশন (দূরদৃষ্টি), সিমুলেশন (অনুকরণ) এবং ডাটা প্রোভেন্সের রিলায়বিলিটি (নির্ভরযোগ্যতা) দিয়ে ডিসিশন-মেকারদের মধ্যে একটা স্পষ্ট উপলব্ধি সৃষ্টি করতে না পারেন, তখন আমাদের ডাটা বিশ্লেষণ-সংশ্লেষণের চর্চা পুরোপুরি ব্যর্থ হয়ে যাবে। অতএব, সতর্কভাবে সমস্যা-সম্পর্কিত ইনফরমেশন কনটেন্টকে সর্বোচ্চ ভিজুয়লাইজেশন করতে হবে।

প্রয়োজনীয় মাল্টিপল টুল ও টেকনিক

দ্রুত বিকশিত প্রায়ুক্তিক পরিবেশে এন্টারপ্রাইজ স্তরের সমস্যার ক্ষেত্রে স্থিতিশীল ও বাড়তি চাপ সহ্য করার মতো (স্ট্যাবল ও স্কেলেবল) টুল ব্যবহারের ব্যাপারে আমাদের সতর্ক হওয়া দরকার। এক্সট্রাক্ট-ট্রান্সফর্ম-লোডের জন্য টুল

(ইনফরমেশন, ওডিআই, গোড়া থেকে), স্থিতিশীলতার জন্য যেমন ডিস্ট্রিবিউটেড কমপিউটিং/প্যারালল কমপিউটিং, ম্যাপ রিডিউস, স্ট্রিমিং, ডাটা প্রসেসিং (অ্যাপাচি স্পার্ক, অ্যাপাচি স্টর্ম), মেশিন-লার্নিং, স্ট্যাটিস্টিকস, ম্যাথামেটিক্যাল মডেলিং অ্যান্ড সিমুলেশন (এসএএস, আর, ম্যাটল্যাব, ম্যাহাউট, এমএল লিব, এসএএস, জেএমপি, মিনিট্যাব, এসপিএসএস, ম্যাথমেটিকা), আর্টিফিশিয়াল ইন্টেলিজেন্স (স্পিচ রিকগনিশন, গুগল স্পিচ এপিআই, মাইক্রোসফট স্পিচ এপিআই, নুয়েস এএসআর), ইন্টেলিজেন্ট কনটেন্ট-অ্যাওয়ার ন্যাচারাল ইন্টারফেসেস, ন্যাচারাল ল্যাঙ্গুয়েজ প্রসেসিং (জিএটিই, অ্যাপাচি ওপেন এনএলপি, এনএলটিকে, স্ট্যানফোর্ড পার্সার, ট্যানসরফ্লোহাস সিনটেক্সনেট), অপারেশন রিসার্চ (আইবিএম আইলগ সিপিএলএক্স, অপটিমাইজেশনের জন্য) এবং ভিজুয়লাইজেশন।

এ ছাড়া অ্যালগরিদমের পাশাপাশি ছোট-বড় ডাটাসেটের জন্য পরীক্ষণেরও প্রয়োজন রয়েছে।

কিছু প্রশ্ন নিয়ে আরও আলোচনা দরকার

এখানে এমন কতগুলো প্রশ্ন উল্লেখ করা হলো, যেগুলো নিয়ে আরও আলোচনা ও বিতর্ক প্রয়োজন।

০১. আমরা কি ভবিষ্যৎ ঘটনা সম্পর্কে ভবিষ্যদ্বাণী করতে পারি (যেমন-প্রতারণা/বিপর্যয়/ঝুঁকি)? যদি তা পারা যায়, তবে কীভাবে? আমরা কি অধিকতর ভালো সিদ্ধান্ত নেয়ার ক্ষেত্রে একটি ডোমেইনের ব্যাপারে আমাদের পদার্থবিদ্যার প্রাতিষ্ঠানিক উপজ্ঞান একীভূত করতে পারি, যার অন্তর্ভুক্তান আনা হয়েছে কোনো ইন্টেলিজেন্ট সিস্টেম থেকে?

এ পর্যন্ত আমরা দেখেছি, যেসব মডেলে রয়েছে গভীরতর পদার্থবিদ্যার বোঝাপড়া, সেখানে সাফল্যের হারটা থাকে বেশি। আর এখানে প্রশ্ন তোলা যাবে ব্ল্যাক-বক্স মডেলের তুলনায় বেশি সহজভাবে। যেমন- ইমেজ রিকগনিশনের জন্য ডিপ ন্যাচারাল নেটওয়ার্ক।

০২. স্বাধীন চলকের সীমানার ভেতরে থেকে আমরা কি উচ্চতর অর্থনৈতিক উপকার পাওয়ার জন্য পর্যাপ্ত পরিমাণে যথাযথভাবে আমাদের 'ওয়ার্ল্ড'কে সিমুলেট ও অপটিমাইজ করতে পারি পথনির্দেশ বা একটি প্রেসক্রিপশন পাওয়ার জন্য?

ডাটা অ্যানালাইসিস এই প্রেসক্রিপশনের জন্য যথেষ্ট নয়। এখানে মডেলিং, সিমুলেশন ও অপটিমাইজেশনের মাধ্যমে সংশ্লেষণ বা সিনথেসিসের প্রয়োজন হয়। এখানে অ্যাকাডেমিয়া ও ইন্ডাস্ট্রিগুলোর উদাহরণ জানলে ভালো হয়।

০৩. নির্ভরযোগ্যভাবে অটোমেট বা সেমি-অটোমেট জেনারেশন ও ভিজুয়লাইজেশন করতে কী প্রয়োজন হয়? আমরা কি লোকজনকে তাদের ন্যাচারাল ল্যাঙ্গুয়েজ (স্পিচ ও টেক্সট) ব্যবহার করে তাদের এন্টারপ্রাইজ ডাটা সোর্সে প্রবেশ করতে দিতে পারি? সিকিউরিটির বিষয়টিই বা কী? সম্ভবত বিষয়টি ইন্ডাস্ট্রির সামান্য কয়টি উদাহরণের চেয়ে বেশি কিছু। এ প্রশ্নে জানা দরকার ইন্ডাস্ট্রির আরও সাফল্য-কাহিনী।

ডাটা সায়েন্স : একটি আকর্ষণীয় ক্যারিয়ার

'হার্ভার্ড বিজনেস রিভিউ' ডাটা সায়েন্সকে অভিহিত করেছে একশতম শতাব্দীর 'সেক্সিয়েস্ট ক্যারিয়ার' অভিধায়। এই কর্মক্ষেত্রে ডাটা স্কলন ও ব্যাখ্যা করার জন্য প্রয়োজন হয় নানা টুল। এই ডাটা স্কলন ও ব্যাখ্যা দেয়া হয় সাধারণত কোম্পানিগুলোকে সহায়তা করতে, যাতে কোম্পানি ভেতরের সবকিছু ভালোভাবে জেনে ভবিষ্যদ্বাণী করতে পারে। গ্লান্সডোর অনুসারে, একজন ডাটা বিজ্ঞানীর গড় বেতন ১১৩,৪৩৬ ডলার। আপনি কী করে এই আকর্ষণীয় ক্যারিয়ারটি আপনার কজায় আনতে পারবেন? নিচে সে সম্পর্কে রয়েছে কিছু পরামর্শ।

০১. আপনি যদি এখনও কলেজেই পড়াশোনা করেন, তবে একটি শক্ত ভিত্তি গড়ে তুলুন প্রবাবিলিটি, স্ট্যাটিস্টিকস ও জেনারেল প্রোথ্রামিংয়ের ওপর। এখন বিশ্ববিদ্যালয়গুলো সবেমাত্র শুরু করছে স্নাতক-পূর্ব ছাত্রদের জন্য ডাটা সায়েন্স প্রোথ্রাম। এ ব্যাপারে কিছু কিছু কলেজ কোর্সও রয়েছে, যেগুলো সম্পন্ন করে নিজেকে তৈরি করতে পারেন ডাটা বিজ্ঞানী হওয়ার ব্যাপারে। 'ডাটা সায়েন্স ক্যারিয়ার গড়ে তুলতে হলে আপনার প্রয়োজন প্রবাবিলিটি ও স্ট্যাটিস্টিকস সম্পর্কে মৌলিক জ্ঞান এবং একই সাথে প্রয়োজন জেনারেল পরপাস প্রোথ্রামিং'- এ অভিমত শাম মোস্তাফার। তিনি 'কোরিলেশনওয়ান'-এর প্রতিষ্ঠাতা ও প্রধান নির্বাহী। কোরিলেশনওয়ান হচ্ছে একটি অনলাইন সার্ভিস, যেখানে ডাটা বিজ্ঞানীরা সুযোগ পান নিয়োগ দাতাদের সাথে মানিয়ে নেয়ার। শাম মোস্তাফা আরও বলেন, যদি কলেজে ডাটা সায়েন্স প্রোথ্রামের সুযোগ না দেয়, তবে ছাত্রদের প্রধান বিষয় তথা মেজর সাবজেক্ট হিসেবে পড়তে হবে কমপিউটার সায়েন্স এবং এর সাথে থাকবে মেজর বা মাইনর সাবজেক্ট হিসেবে স্ট্যাটিস্টিকস।

০২. শানিয়ে নিন আপনার কমিউনিকেশন স্কিল : 'কোর্সওয়ার্কের বাইরে মেজর হিসেবে না হলেও কলেজ ছাত্রদের দলগতভাবে করতে হবে হ্যান্ডসঅন ডাটা অ্যানালাইসিস প্রজেক্ট। ক্রসফাংশনাল টিমের কাজ ছাত্রদের জন্য কমিউনিকেশন স্কিল গড়ে তুলতে সহায়ক হয়। আর এটি হচ্ছে ডাটা সায়েন্সের ক্ষেত্রে একমাত্র গুরুত্বপূর্ণ দক্ষতা'- এ পরামর্শ শাম মোস্তাফার।

০৩. পাইথনের মতো শিখুন স্ক্রিপ্টিং ল্যাঙ্গুয়েজ : টেকনিক্যাল ও নন-টেকনিক্যাল ছাত্ররা সমভাবে উপকৃত হতে পারে পাইথনের মতো একটি স্ক্রিপ্টিং ল্যাঙ্গুয়েজ শিখে নিয়ে। শাম মোস্তাফা বলেন, 'গুগলের রয়েছে একটি সলিড পাইথন অ্যাপ্লাইড কোর্স। এ ছাড়া বেশ কিছু আকর্ষণীয় পাইথন প্রজেক্ট রয়েছে, যেগুলো অনলাইনে পাওয়া যায়।'

'আমি পাইথন প্রোথ্রামিং প্রয়োগের ক্ষেত্রে যত বাস্তব অভিজ্ঞতা বা দক্ষতা অর্জন করেছি, এর সবটুকুই করেছি নিজে নিজে'- জানান বিকামিং অ্যাডাটাসায়েন্টিস্ট ডটকমের ডাটা অ্যানালিস্ট রেন টিয়েটি।

০৪. জ্ঞান অর্জন করুন সে বিষয়ে, যে বিষয়ে (বাকি অংশ ৩১ পৃষ্ঠায়)

হতে চান বিশেষজ্ঞ : ডাটা বিজ্ঞানীরা কাজ করতে পারেন বিভিন্ন ধরনের ইন্ডাস্ট্রিতে— হেলথকেয়ার, ফিন্যান্স, এনার্জি, ট্রান্সপোর্টেশন, এবং আরও অনেক। আপনি যে ইন্ডাস্ট্রিতে কাজ করতে আগ্রহী, সে ইন্ডাস্ট্রির ভেতর-বাইর ভালো করে জানুন, অর্জন করুন ডোমেইন নলেজ। শাম মোস্তাফা বলেন— ‘একজন বড়মাপের ডাটা বিজ্ঞানীর থাকে ব্যাপকভিত্তিক ডোমেইন নলেজ। কখনও কখনও ডোমেইন নলেজ সহায়ক হয় উন্নততর প্রিডিক্টিভ মডেল তৈরি করতে। যথার্থ সঠিকভাবে ডাটা ব্যাখ্যা করতে ডোমেইন নলেজ সহায়ক।’

০৫. ডাটা সায়েন্স ইমারসিভ প্রোগ্রামে যোগ দেয়ার বিষয়টি বিবেচনা রাখুন : আমাদের সবার মাঝে কলেজে ফিরে যাওয়া কিংবা মাস্টার্স ডিগ্রি অর্জন করার মতো বিলাসিতা কাজ করে না। ডাটা সায়েন্সে ট্রানজিশনের একটি উপায় হচ্ছে, ইমারসিভ প্রোগ্রামে যোগ দেয়া। ইমারসিভ প্রোগ্রাম হচ্ছে দ্বিতীয় আরেকটি ল্যান্ডমার্ক শেখার প্রোগ্রাম।

ডাটা সায়েন্স শুধু ডাটা বিজ্ঞানীদের জন্য নয়

ছবির ক্যাপশন : বাজার সচেতনতা ও স্মার্ট

ডাটা ডিসকভারির অ্যাডাপশন ব্যাপক ধরনের ব্যবহারকারীর কাছে ডাটা সায়েন্সকে সম্প্রসারিত করবে, বাড়িয়ে তুলবে অ্যানালাইটিকের প্রভাব—ছবি : ক্লিয়ারস্টোরি ডাটা

আজকে আমরা জানি, ডাটা সায়েন্স শুধু ডাটা সায়েন্টিস্টদের জন্য নয়। ইন্টারনেট অব থিংস, বিগ ডাটা অ্যানালাইটিক ও মেশিন-লার্নিংয়ে আমাদের ক্রমবর্ধমান ব্যবসায়-কেন্দ্রিক উদ্যোগে কমপক্ষে এটিই হচ্ছে দৃশ্যমান প্রবণতা। নতুন ডাটা-অ্যাওয়ার বোর্ডরুম দেখছে এমনকি সিইও ও সিএফওরাও ডাটা হেলথের ব্যাপারেও চাইছেন আইটি ফাংশন। গড়পড়তা ওয়ার্ল্ড ফিউচারিস্ট অথবা সোসাইটি-ডেভেলপমেন্ট কনফারেন্সের দিকে নজর দিন, এটা দেখা অস্বাভাবিক নয় যে, সেখানে অধিবেশন হচ্ছে আর্টিফিশিয়াল ইন্টেলিজেন্স, রোবটিকস ও ডাটা লার্নিং নিয়ে।

০৬. ডাটা রেংলিং, কাউবয় স্টাইল : ব্যবসায়ীরা এখন জানতে চান, তাদের প্রতিষ্ঠান কতটুকু ভালোভাবে ডাটা প্রসেস করছে। এমনকি যদিও এরা জানতে চায় না, এই টুলের পেছনের মেকানিকস এএলএল (অ্যাসোসিয়েশন অব ল্যান্ডমার্ক লার্নিং) সম্পর্কে। মোটের ওপর এ ক্ষেত্রে কার্যকর প্রযুক্তির (ফাংশনাল টেকনোলজির) মধ্যে বেশিরভাগই হচ্ছে অটোমেশন সফটওয়্যারের বা টুলের খণ্ডাংশ। এই

ডাটা অ্যানালাইসিসের উদ্যোগে কোনো না কোনোভাবে বর্ণনা করা হয় সেই পদ্ধতি, যা প্রয়োগ করা হয় ‘ক্লিয়ারস্টোরি ডাটা’য়। এই প্রতিষ্ঠান বলে তথাকথিত ডাটা ইনফারেন্সের কথা এবং এটি কোনো না কোনো উপায়ে ‘ইন্টেলিজেন্ট ডাটা হারমোনাইজেশন’ পদবাচ্যটির ট্রেডমার্ক তৈরি করতে সক্ষম হয়েছে। এখানে আমরা ‘ইনফাইনিট ডাটা ওভারলেপ ডিটেকশন’ নামের আরেক টুকরো টেকনোলজি পাই, যার অভ্যন্তরীণ কোডনেম হচ্ছে আইডিওডি। এটি একটি স্পার্কভিত্তিক অ্যানালাইটিকস প্রোডাক্ট, যা প্রতিটি সোর্সে ডাটা প্যাটার্ন ও কাস্টমার-স্পেসিফিক ডাটা টাইপ চিহ্নিত করতে সক্ষম বলে দাবি করে। এখানে একজন ব্যবহারকারী সংশ্লিষ্ট হয় একটি অ্যানালাইসিসের অংশ হিসেবে।