ডেটা এবং মেশিন লার্নিংঃ কোন দিকে যাচ্ছে বিশ্ব?

* ফারহান ইশরাক খান

ishraque.farhan@gmail.com

বর্তমান আইটি বাজারে “বিগ ডেটা”, “ক্লাউড”, “আইওটি”, “মেশিন লার্নিং” – কথাগুলো খুব প্রচলিত, তাই না? কিন্তু এই যে এত এত ডেটার কথা আমারা বলি, তার কতটুকু আসলে আমরা ব্যবহার করছি; আমাদের কতটুকুই বা সামর্থ্য আছে? আমাদের দৈনন্দিন জীবনের সাথেও কিন্তু এখন মেশিন লার্নিং এর মত কঠিন বিষয়গুলো জড়িয়ে গেছে। ফেসবুকে আমরা যখন ছবি আপলোড করি তখন বন্ধুদের ট্যাগ সাজেশন দেয়া হয় মেশিন লার্নিং টেকনিক ব্যবহার করে। মাইক্রোসফট কাইনেকট কিংবা সনি পিএস গেমিং কনসোল এও ব্যবহার করা হয় এই টেকনিক। আবহাওয়ার পূর্বাভাসেও দিন দিন বৃদ্ধি পাচ্ছে মেশিন লার্নিং এর ব্যবহার। গুগল কিংবা ফেসবুক আপনাকে যে আপনাকে পছন্দের বিষয়গুলি সামনে এনে দিচ্ছে, এর পেছনেও আছে মেশিন লার্নিং এর ভূমিকা। মেশিন লার্নিং নিয়ে সম্ভাবনা কতটুকু – এই প্রশ্নের উত্তর খোঁজার চেষ্টায় এই লেখার ক্ষুদ্র প্রয়াশ।

**মেশিন লার্নিং কি?**

মেশিন লার্নিং কম্পিউটার বিজ্ঞানের ক্ষেত্রে কৃত্রিম বুদ্ধিমত্তা প্রদানের একটি পদ্ধতি যেখানে সরাসরি কম্পিউটারকে “কি করতে হবে”- তা বলে না দিয়ে কিছু ঘটনাবলির নীরিক্ষে কি করতে হবে তা শেখানো হয়। এ ক্ষেত্রে পরিসংখ্যান এবং কম্পিউটার বিজ্ঞানের অনেক এলগরিদম ব্যবহার করা হয়। ব্যাপারটা অনেকটা বাচ্চাকে ভাল- মন্দের পার্থক্য শেখানোর মত। প্রথমে বাচ্চাদের কোনটা ভাল আর কোনটা মন্দ তা শেখানো হয়, ধীরে ধীরে মানুষ ভাল মন্দের পার্থক্য নিজেই করতে শেখে।

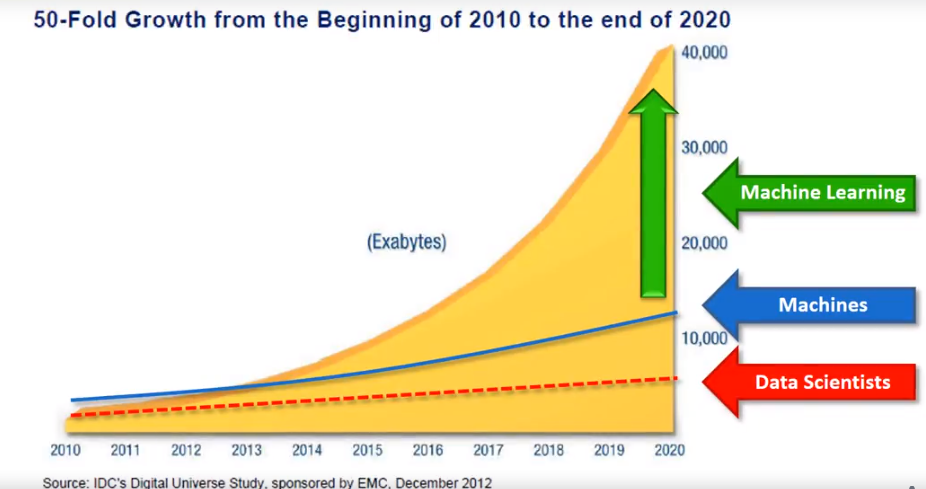
মানুষের বাচ্চাকে শেখানোর যেমন কিছু উপায় আছে, যন্ত্রকে শেখানোরও তেমন কিছু উপায় আছে। কখনও বা যন্ত্রকে উদাহরণ দিয়ে দেয়া হয় – কি করলে কি হবে। মেশিন লার্নিং এর ভাষায় একে বলে সুপারভাইসড লার্নিং (Supervised learning) । আবার অনেক সময় কাজের বিবরণ থাকে কিন্তু ফলাফল বলে দেয়া হয় না। মেশিনকে নিজের হিসেব করে বের করতে হয় (unsupervised learning) । মোদ্দাকথা, মানুষ তার সন্তানদেরকে যেভাবে শেখায় মেশিনকেও সেভাবে শেখানোর চেষ্টা করা হয়। কিন্তু, মেশিনকে এত শিক্ষিত করে লাভ টা কি? চলুন তাহলে দেখে নিই মেশিন লার্নিং এর আসল সম্ভাবনা টা কোথায়?

**তথ্য প্লাবন এবং মেশিন লার্নিং**

বর্তমান সময়ে আমরা তথ্য প্লাবন (data deluge) এর কথা প্রায়ই বলে থাকি। আমাদের দৈনন্দিন জীবনে প্রতিনিয়ত আমরা তথ্য তৈরি করে যাচ্ছি। এক এক জন মানুষ এক একটা তথ্য তৈরির কারখানা। বিভিন্ন প্রতিষ্ঠানও প্রতিনিয়ত তথ্য তৈরি করে যাচ্ছে। বিভিন্ন মাধ্যমে আমরা সেই তথ্যের কিছুটা ছড়িয়ে দিচ্ছি আন্তরজালে। সৃষ্টির শুরু থেকে মানব জাতি আসলে কত টুকু তথ্য তৈরি করেছে?

* সৃষ্টির শুরু থেকে ২০০৫ সাল পর্যন্ত সমগ্র মানবজাতি ১৩০ এক্সাবাইট ( ১ এক্সাবাইট = ১০০০ পেটাবাইট) তথ্য তৈরি করেছিল। এই তথ্যের মাঝে মানুষের তৈরি সকল বই, গান, ছায়াছবি, স্থিরচিত্র সব কিছু অন্তর্ভুক্ত। কিন্তু এক্সাবাইট আসলে কতটুকু? আমাজন জঙ্গলে প্রায় ৭০০ বিলিয়ন গাছ আছে। আপনি যদি অ্যামাজনের সকল গাছ কেটে তা দিয়ে বই বানিয়ে তাতে তথ্য সংগ্রহ করেন তবে আপনি সর্বোচ্চ ১ পেটাবাইট তথ্য সংগ্রহ করতে পারবেন। তাহলে ১৩০ এক্সাবাইট তথ্য লিখে রাখার জন্য আমাদের ১৩০০০০ টি অ্যামাজন জঙ্গলের গাছ দরকার!
* এতো ছিল ২০০৫ সালের কথা। ২০১০ সালে গিয়ে এই পরিমাণটা কততে দাঁড়ায় জানেন? ১২০০ এক্সাবাইট! আর ২০১৫ সালে এর পরিমাণ হয় ৭৯০০ এক্সাবাইট।
* যদি এই প্রবণতা চালু থাকে, তবে ২০২০ সালে গিয়ে আমরা ৪০,৯০০ এক্সাবাইট ডেটা দেখতে পাব

তাহলে দেখা যাচ্ছে ২০১০ থেকে ২০২০ সাল অবধি ডেটার বৃদ্ধি ৫০ গুণ। এই বিশাল ডেটা থেকে কিভাবে মানুষের দৈনন্দিন জীবনকে আরও সহজ করা যায় তাই নিয়ে ডেটা বিশেষজ্ঞদের কাজ। আর এই ডেটাকে আরও ভালোভাবে বুঝতে পাড়ার একটা উপায় হল মেশিন লার্নিং। এব্যাপারে ২০১২ সালে চালনো আইডিসি- র জরিপ থেকে আরও ভাল একটা ধারনা পাওয়া যায়ঃ



* পরিসংখ্যানের গতানুগতিক ধারনাগুলো ব্যবহার করে হয়ত ৬০০০ এক্সাবাইট পর্যন্ত ডেটার রহস্য উদ্ঘাটন করা সম্ভব।
* ২০২০ সালনাগাদ গতানুগতি ডেটা ওয়ারহউসিং, বিজনেস ইন্টেলিজেন্স, বিগ ডেটা টুল ব্যবহার করে ডেটা সাইন্টিস্টগণ হয়ত ১২০০ এক্সাবাইট ডেটার রহস্য উদ্ঘাটন করতে পারবেন
* বাকি সম্পূর্ণ ডেটার রহস্য উদ্ঘাটন করার সম্ভাবনা নির্ভর করে মেশিন লার্নিং পদ্ধতির উন্নয়ন ও ব্যাবহারের উপর

এত কথা বলার উদ্দেশ্য আসলে মেশিন লার্নিং এর মাহাত্ত বর্ণনা নয়, আমাদের দেশে মেশিন লার্নিং নিয়ে কাজ করার সুযোগ খোঁজা। তথ্যপ্রযুক্তি খাতে মেশিন লার্নিং অনেক নতুন কিছু যোগ করেছে এবং আরও করবে। এখন সময় বাংলাদেশের এই খাতে সত্যিকারের কিছু করার।