

# হাতেকলমে মেশিন লার্নিং

পরিচিতি, প্রজেক্ট টাইটানিক, আর এবং পাইথনসহ

২য় সংস্করণ

রকিবুল হাসান

## সূচি

<b>শুরুর আগে : টুকটাক কথা</b>	<b>১১</b>
কাদের জন্য বইটি বেশি জরুরি?	১৩
কোনটা দিয়ে শুরু করব?	১৬
কীভাবে পড়বেন বইটি?	১৮
<b>দ্বিতীয় সংস্করণের কথা</b>	<b>২০</b>
<b>মুখবন্ধ</b>	<b>২২</b>
<b>বদলে যাওয়ার ঘটনা</b>	<b>২৪</b>
১.১ সিডিসির ওয়ার্নিং	২৫
১.২ ইন্টারকানেকশন কস্ট মডেলিং	২৮
১.৩ মানুষের পাশে ডেটা	২৯
১.৪ গুগল ফটোজ	৩৩
১.৫ গুগল বাংলা ‘টেক্সট টু স্পিচ’ আর আইওটি (IoT)	৩৪
১.৬ সরকারি ওপেন ডেটা	৩৪
১.৭ ডেটা শেয়ারিং : প্রাইভেট ও পাবলিক সেক্টর	৩৬
১.৮ আর্টিফিশিয়াল ইন্টেলিজেন্স : মেশিন লার্নিংয়ের প্রসার	৩৯
১.৯ মেশিন লার্নিং ব্যবহারে যে ইন্ডাস্ট্রিগুলো এগিয়ে	৪২
<b>ডেটা থেকে ভবিষ্যৎ দেখার ধারণা</b>	<b>৪৫</b>
২.১ মেশিন লার্নিং শেখার সহজ পথ	৪৬
২.২ চিটকোড	৪৭
২.৩ ভবিষ্যৎ দেখার ধারণা	৪৮
২.৪ ডেটার গল্প বলার ক্ষমতা	৪৯
২.৫ ‘মৃত্যু’ অ্যালগরিদম	৫০
২.৬ দেশের নীতিনির্ধারণী ড্যাশবোর্ড	৫২

<b>মেশিন লার্নিং কী এবং কেন?</b>	<b>৫৫</b>
৩.১ মেশিন লার্নিং কী?	৫৬
৩.২ কেন দরকার মেশিন লার্নিং?	৫৮
৩.৩ মেশিন লার্নিংয়ের কয়েকটি ভাগ	৬১
৩.৪ ডেটা ব্যবহারের শুরুর ধারণা	৬২
৩.৫ মেশিন লার্নিং অ্যালগরিদম	৬৩
৩.৬ মেশিন লার্নিং মডেল	৬৫
৩.৭ ট্রেনিং ডেটা ও টেস্ট ডেটা কী?	৬৬
৩.৮ মেশিন লার্নিং : ভেতরের কিছু কথা	৭০
৩.৯ ডেটার তত্ত্ব এবং প্রসেস	৭১
৩.১০ মেশিন লার্নিং যেভাবে শিখবেন	৭২

<b>ক্যাগল প্রতিযোগিতা</b>	<b>৭৪</b>
৪.১ ক্যাগল কী?	৭৫
৪.২ গুগল কেন ক্যাগল কিনেছে?	৭৬
৪.৩ ক্যাগলে কাজ করার পদ্ধতি	৭৮
৪.৪ কেন প্রজেক্ট দিয়ে শুরু?	৮০
৪.৫ কেন শুরুরতই আর প্রোগ্রামিং এনভায়রনমেন্ট?	৮১
৪.৬ মেশিন লার্নিং কমিউনিটি	৮৪
৪.৭ ক্যাগল কার্নাল এবং অনলাইন হোস্টেড স্ক্রিপ্ট	৮৫

<b>আর এনভায়রনমেন্ট</b>	<b>৮৭</b>
৫.১ আর এবং আরস্টুডিও	৮৮
৫.২ আরস্টুডিও	৯০
৫.৩ প্রজেক্ট টাইটানিক এবং আরস্টুডিও	৯২
৫.৪ গিটহাব থেকে স্ক্রিপ্ট ডাউনলোড	৯৩
৫.৫ প্রজেক্টের গিটহাব স্ক্রিপ্ট	৯৪

<b>প্রজেক্ট টাইটানিক : বিপর্যয় থেকে শেখা</b>	<b>৯৫</b>
৬.১ কেন প্রজেক্ট 'টাইটানিক'?	৯৬
৬.২ টাইটানিক : বিপর্যয় জানতে মেশিন লার্নিং	৯৬
৬.৩ টাইটানিকের ডেটা	৯৯
৬.৪ কেন এই ডেটা সেট?	১০০
৬.৫ ট্রেনিং ডেটা ও টেস্ট ডেটা সেট	১০০
৬.৬ আমাদের টাইটানিক চ্যালেঞ্জ	১০১
৬.৭ প্রজেক্টে ক্যাগলের ব্যবহৃত সব কার্নাল	১০৩
৬.৮ আরস্টুডিওতে ডেটা লোড	১০৫

৬.৯ ফাইল মেন্যু থেকে ডেটা ইমপোর্ট	১০৫
৬.১০ স্ক্রিপ্ট দিয়ে ডেটা ইমপোর্ট	১০৭
<b>মেশিন লার্নিং প্রেডিকশন : টাইটানিক ডেটা সেট</b>	<b>১১০</b>
৭.১ প্রথম প্রেডিকশন	১১১
৭.২ ক্যাগলে সাবমিশনের জন্য প্রস্তুতি	১১৪
৭.৩ সাবমিশন ফাইল তৈরি	১১৫
৭.৪ ক্যাগলে সাবমিশন	১১৫
৭.৫ ডেটা ভিজুয়ালাইজেশন	১১৬
৭.৬ দ্বিতীয় প্রেডিকশন	১২০
৭.৭ তৃতীয় প্রেডিকশন	১২৩
৭.৮ ডিসিশন ট্রি, কাহিনি কী?	১২৯
৭.৯ চতুর্থ প্রেডিকশন : মেশিন লার্নিং গুরু	১৩৩
৭.১০ পঞ্চম প্রেডিকশন (ফিচার ইঞ্জিনিয়ারিং)	১৩৫
৭.১১ ডেটা প্রি-প্রসেসিং, ডেটা ক্লিনিং এবং ফিট প্রেডিকশন	১৫১
৭.১২ সপ্তম প্রেডিকশন (র্যান্ডম ফরেস্ট)	১৫৬
<b>কী আছে সামনে?</b>	<b>১৬৬</b>
৮.১ এ পর্যন্ত কী শিখলাম আমরা?	১৬৭
৮.২ কোথায় যাচ্ছি এরপর?	১৬৯
<b>পাইথন দিয়ে টাইটানিক প্রজেক্ট</b>	<b>১৭০</b>
৯.১ জুপিটার নোটবুক ইনস্টলেশন	১৭১
৯.২ টাইটানিক জাহাজডুবিতে বেঁচে যাওয়ার প্রেডিকশন	১৭২
৯.৩ এক্সপ্লোরেটরি ডেটা অ্যানালাইসিস	১৭৫
৯.৪ ডেটা ভিজুয়ালাইজেশন	১৭৮
৯.৫ ফিচার ইঞ্জিনিয়ারিং থেকে মডেলের অ্যাকুরেসি বাড়ানো	১৮২
৯.৮ মেশিন লার্নিং মডেলিং	২০১
৯.৯ ক্যাগলে আপলোড	২০৩
<b>পরিশিষ্ট</b>	<b>২০৫</b>
সামনের বই	২০৬
যোগাযোগের মাধ্যম	২০৬

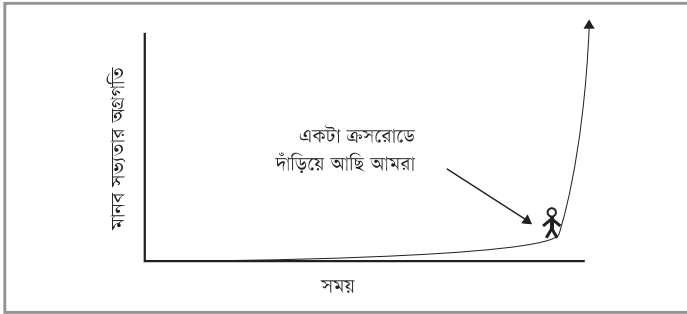
## ১.৮ আর্টিফিশিয়াল ইন্টেলিজেন্স : মেশিন লার্নিংয়ের প্রসার

*Just as eating contrary to the inclination is injurious to the health, so study without desire spoils the memory, and it retains nothing that it takes in. — Leonardo da Vinci*

আর্টিফিশিয়াল ইন্টেলিজেন্স মানে কৃত্রিম বুদ্ধিমত্তার একটা ছোট্ট অংশ হচ্ছে মেশিন লার্নিং। একটা সাবসেট। কৃত্রিম বুদ্ধিমত্তার যে অংশটুকু যন্ত্রকে কানেক্ট করতে হয়, সেখানে চলে আসে ওই মেশিনকে শেখানোর কাজ। মেশিনকে শেখাব কীভাবে? আগেই বলেছি, তার অভিজ্ঞতা দিয়ে। আর তার অভিজ্ঞতা হচ্ছে পুরোনো, মানে ‘লিগেসি’ ডেটা।

এটা ঠিক যে একটা ক্রসরোডে দাঁড়িয়ে আমরা। সামনের ২০ বছরে যা উন্নতি হবে, তা এই মুহূর্তে ভাবতেই পারছি না। চতুর্থ ইন্ডাস্ট্রিয়াল রেভল্যুশন নিয়ে আসবে ‘আর্টিফিশিয়াল ইন্টেলিজেন্স’। পৃথিবীতে মানুষ আছে ২.৮ মিলিয়ন বছর আগে থেকে। ওই সময় থেকে আজ পর্যন্ত যা উন্নতি হয়েছে, তার দ্বিগুণেরও বেশি হবে এই ২০ বছরে!

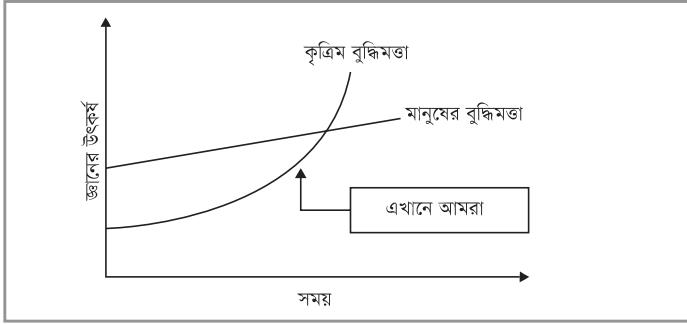
মানুষের মতো জিনিসপত্র আনা-নেওয়া, ওঠানো-নামানো—এসব পারে যন্ত্র। বুদ্ধিমত্তায় এখনো কিছুটা পাখির মতো। বর্তমান গতিতে এগোলে ২০৩৫-৪০-এর মধ্যে মানুষের মতো বুদ্ধি ধরবে যন্ত্র।



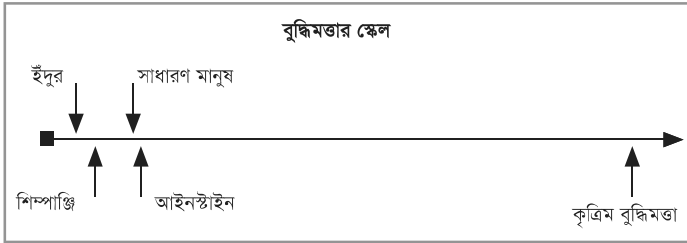
চিত্র ১.২ : মানব সভ্যতার একটা ক্রসরোডে দাঁড়িয়ে আছি আমরা

আমার ধারণা, জ্ঞানের ‘ডিফিউশন’ আসে প্রিন্টিং প্রেস আবিষ্কারের পর থেকে। ১৪৪০-এর পর থেমে থাকেনি সময়। বাষ্পীয় ইঞ্জিন মানুষকে দেয় কৃত্রিম শক্তি, অনেক অনেক বেশি হর্স পাওয়ার। তার

আগে ওই ঘোড়াই ছিল ভরসা। এল তেলের ইঞ্জিন। কৃত্রিম সূর্যের মতো ক্ষমতা নিয়ে এল বিদ্যুৎ। পেছনে তাকায়নি মানুষ। ইন্ডাস্ট্রিয়াল রেভল্যুশনে মানুষ এগিয়েছে যন্ত্র দিয়ে। তবে জ্ঞানে নয়। প্রযুক্তি বাড়ে ‘এক্সপোনেনশিয়াল’, মানে গুণীতক গতিতে, বাকি সব গতিতে ‘লিনিয়ার’। যেমন শস্য উৎপাদনে। ম্যুরের তত্ত্বের হিসাবে প্রতি ১৮ থেকে ২৪ মাসে দ্বিগুণ হয় যন্ত্রের ক্ষমতা। এখন ম্যুরের তত্ত্বও ফেল। সেটা কমে এসেছে অনেক। অচিরেই এমন দিন আসছে, যখন মানুষের মতো করেই মানুষকে সাহায্য করবে যন্ত্র।



চিত্র ১.৩ : মানুষের পাশাপাশি কৃত্রিম বুদ্ধিমত্তার উত্থান



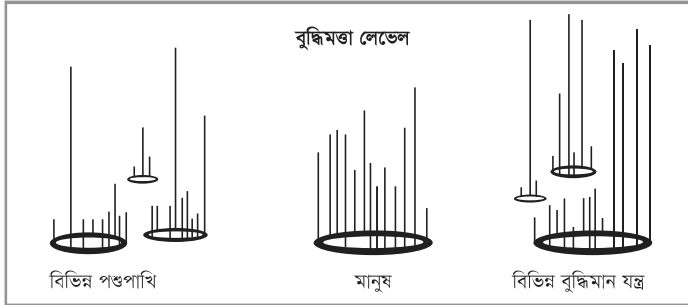
চিত্র ১.৪ : মানুষ, বিভিন্ন প্রাণী ও কৃত্রিম বুদ্ধিমত্তার স্কেল

সবার বুদ্ধিমত্তাকে স্কেলে ফেললে কী হয় দেখি (চিত্র ১.৪)। আমাদের ধারণা, পশুপাখির বুদ্ধিমত্তা এখনো প্রিমিটিভ লেভেলে। শিম্পাঞ্জির বুদ্ধি আমাদের কিছুটা কাছাকাছি। তারপর আমরা, মানে আমজনতা। তারপর আইনস্টাইন অথবা ওই লেভেলের বিজ্ঞানীরা। আসলে কি তা-ই? মানুষের বুদ্ধিমত্তা কি ওখানেই আটকানো? আমাদের অনেকেই

অনেক স্টিটস্মার্ট। অনেকে অনেক কিছু মনে রাখতে পারেন দীর্ঘদিন। কেউ কেউ ইমোশনাল ইন্টেলিজেন্ট। আমার এক বন্ধু আছে যে কার্যকারণে ওস্তাদ। আরেকজন তুখোড় সিদ্ধান্ত গ্রহণে। একেবারে বসম্যান! তাহলে কী হবে এখন?

আমাদের বুদ্ধিমত্তার স্কেলেই আছে বড় ধরনের ঘাপলা। বুদ্ধিমত্তার স্কেল কিন্তু এক ডাইমেনশনের ব্যাপার নয়। অনেকটা গানের নোটের মতো। হাজারো ইনস্ট্রুমেন্ট, গলা, ব্যাকআপ ভোকালিস্ট না থাকলে হয় না পুরো গান। তেমনি দরকারি সব জ্ঞান না থাকলে বুদ্ধিমত্তা অপূর্ণ।

আমরা বিজ্ঞানে আইনস্টাইন না হতে পারি কিন্তু ‘ইমোশনাল ইন্টেলিজেন্স’ লেভেলে তো কম নই। ধরে নিন, আমি কিছু কিছু জিনিসে ভালো হব অন্য অনেকের চেয়ে। আবার অন্যেরা আমার চেয়ে ভালো অন্য কিছু বিষয়ে। একেক মানুষের একেক দিকে জ্ঞানের লেভেল বেশি। মানুষের মাথা কিন্তু ভর্তি হাজারো জিনিস দিয়ে।



চিত্র ১.৫ : বিভিন্ন প্রাণীর বুদ্ধিমত্তার লেভেল (উৎস : কেভিন কেলি)

মানুষ যে সবকিছুতেই স্মার্ট, সেটাই বা বলল কে? একটি কাঠবিড়াল কয়টা বাদাম পুরো মাঠের কোথায় কোথায় রেখেছে, সেটা মানুষের চেয়ে ভালো মনে রাখতে পারে। হাতি তো ভোলে না কিছুই। আবার কিছু মানুষের স্মৃতি গোল্ডফিশের চেয়েও খারাপ। কিছু কুকুর তার মনিবকে চেনে তার বউয়ের থেকে বেশি। আবার পশুপাখি আগেভাগেই বলতে পারে প্রাকৃতিক বিপদের কথা—যা মানুষের পক্ষে অসম্ভব।

সামান্য ক্যালকুলেটর তার বুদ্ধিমত্তায় মানুষকে ছাড়িয়েছে অনেক দিন আগেই। কিন্তু সেটা তো শুধু অঙ্কে। জিপিএসের বুদ্ধিমত্তা ভালো নেভিগেশনে। গুগল মানুষ থেকে অনেক ভালো—লং টার্ম মেমোরিতে। ১.৫ নং চিত্রটা দেখুন। একেকটা যন্ত্র একেক ফিল্ডে ভালো। কিন্তু মানুষের বুদ্ধিমত্তা অনেক অনেক ডাইমেনশনের।

তাহলে যন্ত্র কবে সব দিক থেকে মানুষের চেয়ে বুদ্ধিমান হবে? মনে আছে ইলেকট্রিক গ্রিডের কথা? আপনার বাসা ততটুকুই নেয়, যতটুকু দরকার। কৃত্রিম বুদ্ধিমত্তাও আসবে গ্রিড দিয়ে। মানুষের থেকেও বেশি বুদ্ধিমত্তা নিয়ে। আপনি নেবেন যতটুকু দরকার, ততটুকু।

পানি তোলার জন্য আগে বাসায় ছিল চাপকল। পরে এল কৃত্রিম শক্তি বিদ্যুৎ। গ্রিড থেকে। চাপকল হয়ে গেল পাম্প। এরপর আসবে কৃত্রিম বুদ্ধিমত্তা। ওই গ্রিড থেকে। পাম্প হবে স্মার্ট পাম্প। মানুষের ইশারা ছাড়াই দরকারমতো তুলবে পানি। অফিস থেকে আসার আগেই গোসল অথবা কফির জন্য করবে গরম পানি। কৃত্রিম বুদ্ধিমত্তা গ্রিড কিছু আছে এখনই। ‘ক্লাউড’ নাম তার। গ্যাস, বিদ্যুতের মতো আমরা কিনব কৃত্রিম বুদ্ধিমত্তা। যেভাবে কিনছি গুগলের স্পিচ টু টেক্সট বাংলা অ্যাপ্লিকেশন প্রোগ্রামিং ইন্টারফেস। প্রতি ১৫ সেকেন্ড করে। কথা বলি বাসায়—ভয়েস সিনথেসিস হচ্ছে কোনো না কোনো সার্ভারে। পাওয়ার গ্রিডের মতো কৃত্রিম বুদ্ধিমত্তা আসছে গ্রিড থেকে, ইন্টারনেট হয়ে।

## ১.৯ মেশিন লার্নিং ব্যবহারে যে ইন্ডাস্ট্রিগুলো এগিয়ে

*If we have data, let's look at data. If all we have are opinions, let's go with mine. —Jim Barksdale*

অনেকেই প্রশ্ন করেন, ‘আর্টিফিশিয়াল ইন্টেলিজেন্স’ের প্রসার বেশি কোন ‘ডোমেইনে’? অনেকের ভাষায়—এআই থেকে লাভ পাচ্ছে কোন ইন্ডাস্ট্রি? অথবা এই সময়ে এআইয়ের সুবিধা নিচ্ছে কোন পার্টি?

‘কমিউনিকেশন সিকিউরিটি’ নিয়ে কাজ করি বলে এই জায়গায় এআইয়ের প্রসার চোখে পড়ে বেশি। আবার পিছিয়ে নেই কিন্তু গেমস