

# Session 5&6

## Introduction to Machine Learning (ML)

NLP | Machine Learning | Zahra Amini

Telegram: @zahraamini\_ai & Instagram:@zahraamini\_ai & LinkedIn: @zahraamini-ai

<https://zil.ink/zahraamini>

هوش مصنوعی علم و مهندسی ساخت ماشین‌های هوشمند است، به‌ویژه ماشین‌هایی که می‌توانند کارهایی را انجام دهند که معمولاً به هوش انسانی نیاز دارند.

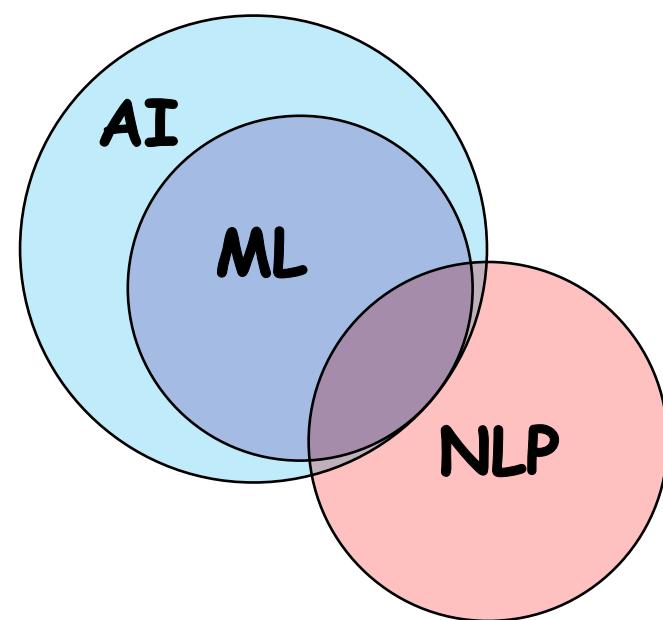
هوش مصنوعی  
Artificial intelligence  
(AI)  
چیست؟

جان مک‌کارتی

به هر تکنیکی که کامپیوترها را قادر می‌سازد تا همانند انسان رفتار کنند.

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به سیستم‌ها و الگوریتم‌ها این امکان را می‌دهد که به صورت خودکار از داده‌ها یاد بگیرند و بدون اینکه به‌طور صریح برنامه‌ریزی شده باشند، عملکرد خود را بهبود بخشدند. در یادگیری ماشین، مدل‌ها از داده‌ها و تجربیات گذشته استفاده می‌کنند تا الگوهای را شناسایی کنند و برای پیش‌بینی یا تصمیم‌گیری در شرایط جدید، خود را تطبیق دهند.

یادگیری ماشین  
Machine Learning  
چیست؟



یک مثال بزنید که یک سیستم هوشمند باشد اما از یادگیری ماشین استفاده نکند

## مثال

## سناریو ۱: سیستم گرمایشی هوشمند با ترموستات ساده

فرض کنید شما یک سیستم گرمایشی دارید که توسط یک ترموستات ساده کنترل می‌شود. این ترموستات به گونه‌ای تنظیم شده که وقتی دمای اتاق از یک حد مشخص پایین‌تر می‌آید، سیستم گرمایشی را روشن می‌کند و وقتی دما به حد بالایی می‌رسد، سیستم را خاموش می‌کند. این ترموستات بر اساس قوانین از پیش تعیین شده عمل می‌کند و هیچ‌گونه یادگیری یا بهبود در عملکرد آن اتفاق نمی‌افتد.

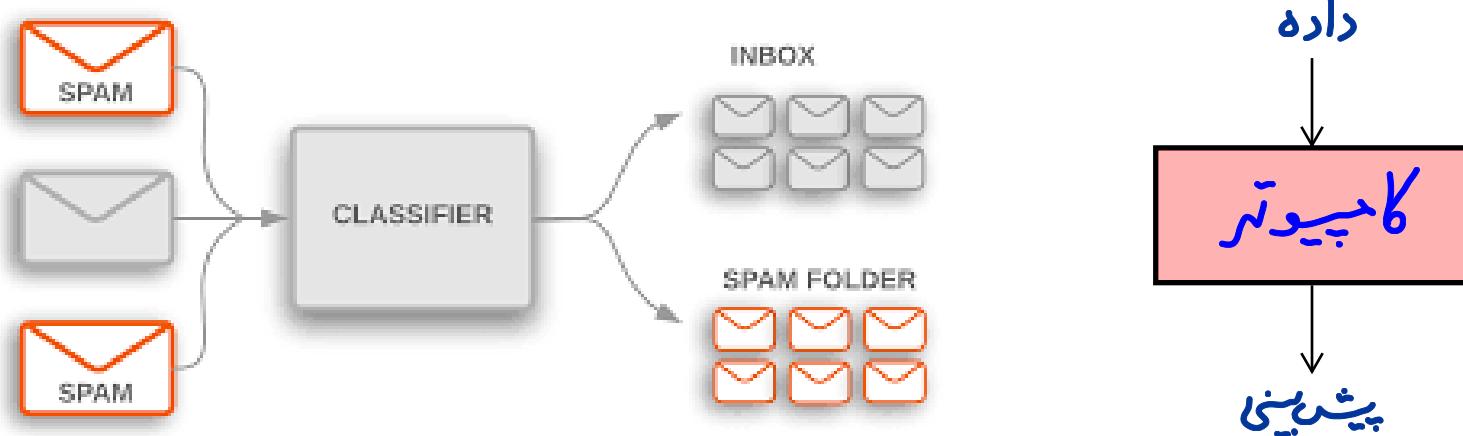
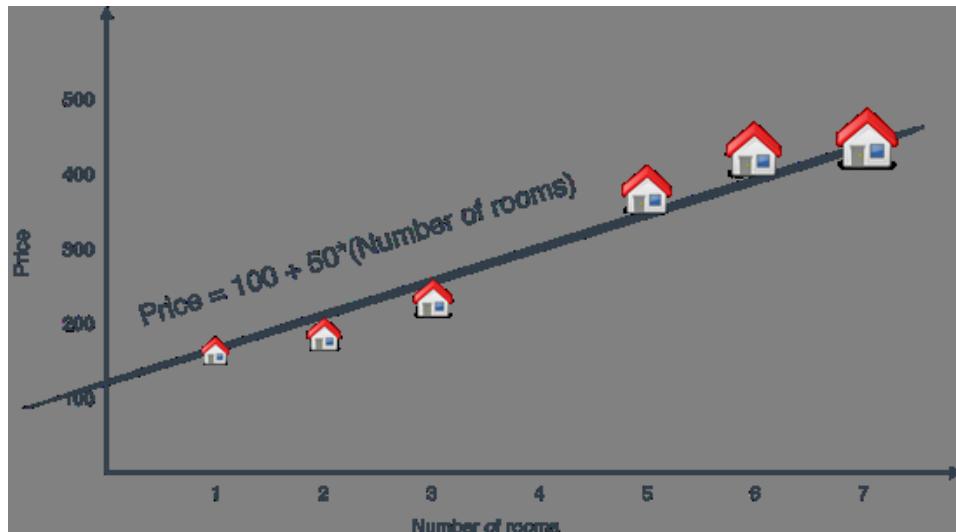
این سیستم هوشمند است زیرا می‌تواند وضعیت دمای اتاق را تشخیص دهد و بر اساس آن تصمیم بگیرد که سیستم گرمایشی را روشن یا خاموش کند. با این حال، این سیستم از یادگیری ماشین استفاده نمی‌کند، چرا که فقط به تنظیمات از پیش تعریف شده متکی است و نمی‌تواند از داده‌های گذشته برای بهبود عملکردش استفاده کند.

## سناریو ۲: سیستم گرمایشی هوشمند با یادگیری ماشین

فرض کنید یک سیستم گرمایشی پیشرفته دارید که با استفاده از یادگیری ماشین و حسگرهای مختلف کار می‌کند. این سیستم علاوه بر اندازه‌گیری دما، اطلاعات دیگری مانند الگوی حضور افراد در خانه، تغییرات فصلی دما، وضعیت آب و هوا، و ترجیحات گرمایشی ساکنین را جمع‌آوری و تحلیل می‌کند. سیستم با پردازش این داده‌ها، می‌تواند به شکلی بهینه و منطقی Rational عمل کند.

به عنوان مثال، سیستم یاد می‌گیرد که در روزهای هفته، صبح‌ها قبل از اینکه ساکنین بیدار شوند، دمای اتاق را به سطح مطلوب افزایش دهد و در ساعت‌های خالی بودن خانه، دما را کاهش دهد. همچنین، در صورت پیش‌بینی هوای گرم در طول روز، سیستم می‌تواند به طور منطقی تصمیم بگیرد که دما را از همان ابتدا خیلی بالا نبرد تا انرژی هدر نزود.

به این ترتیب، این سیستم گرمایشی هوشمند، علاوه بر اینکه به صورت مستقل عمل می‌کند، از یادگیری ماشین بهره می‌برد تا عملکردی بهینه و منطقی داشته باشد. این توانایی یادگیری و بهینه‌سازی، آن را از یک سیستم ساده هوش مصنوعی که تنها به دستورات ثابت متکی است، متمایز می‌کند.



## کاربردهای ML

دیتاست یا مجموعه داده، به مجموعه‌ای از داده‌ها گفته می‌شود که معمولاً در قالب یک جدول یا ماتریس سازماندهی شده‌اند. یک دیتاست معمولاً شامل چندین رکورد یا نمونه است که هر کدام از آن‌ها دارای تعدادی ویژگی یا متغیر می‌باشد. هر ستون دیتاست نشان‌دهنده یک ویژگی خاص است و هر سطر نشان‌دهنده یک رکورد منحصر به فرد است. به عنوان مثال، یک دیتاست مربوط به اطلاعات دانشجویان ممکن است شاملاً ستون‌های زیر باشد:

## دیتاست چیست؟

## مثال

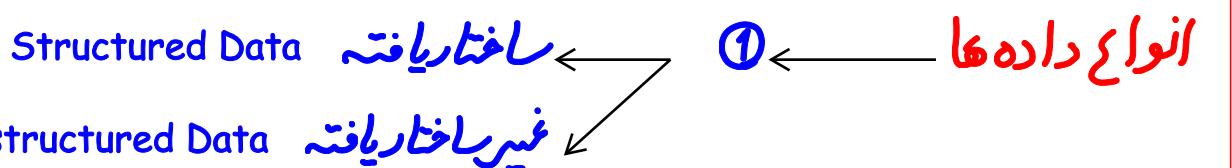
Name	Age	Field of Study	GPA
Ali Rezaei	21	Computer Engineering	17.5
Maryam Ahmadi	23	Medical Sciences	18.2
Hossein Mousavi	22	Civil Engineering	16.8
Sara Alavi	20	Mathematics	19.1
Mohammad Nazari	24	Physics	17.0

نام و نام خانوادگی  
سن  
رشته تحصیلی  
معدل کل

هر سطر از این دیتاست نشان‌دهنده اطلاعات یک دانشجو است.

دیتاست‌ها بسته به نوع داده‌ای که در آن استفاده می‌شوند، می‌توانند انواع مختلفی داشته باشند، مانند دیتاست‌های ساختاریافته (structured) و دیتاست‌های غیرساختاریافته (unstructured).

	Username	Processed Review	and	but	deliveri	did	excel	در این ماتریس، سطراها نمایانگر اسناد (در اینجا، هر نظر یا جمله) و ستون‌ها نمایانگر واژه‌ها یا عبارات کلیدی هستند. هر مقدار در این ماتریس نشان‌دهنده حضور یا اهمیت یک واژه در یک سند خاص است	ساختار ماتریسی دیتاست‌های متنی
1	Ali Rezaei	veri satisfi with the product qualiti	0.0	0.0	0.0	0.0	0.0		
2	Maryam Ahmadi	deliveri wa late but the product wa good	0.0	0.363 7828 0673 4642	0.363 78280 6734 6423	0.0	0.0		
3	Hossein Mousavi	poor qualiti did not meet my expect	0.0	0.0	0.0	0.4 04 90	0.0		



داده‌های ساختاریافته به داده‌هایی گفته می‌شود که دارای قالب و ساختار مشخصی هستند. این داده‌ها معمولاً در قالب جداول و به صورت ردیف و ستون سازمان‌دهی می‌شوند، که هر ستون نشان‌دهنده یک ویژگی خاص و هر ردیف نمایانگر یک رکورد است. داده‌های ساختاریافته معمولاً در پایگاه‌های داده رابطه‌ای (مثل SQL) ذخیره و مدیریت می‌شوند

داده‌های غیرساختاریافته به داده‌هایی گفته می‌شود که فاقد قالب یا ساختار مشخصی هستند و نمی‌توان آن‌ها را به راحتی در قالب جداول سازمان‌دهی کرد. این نوع داده‌ها معمولاً شامل متن‌های طولانی، تصاویر، ویدیوها و صداها هستند و به راحتی در پایگاه داده‌ای رابطه‌ای ذخیره نمی‌شوند

## أنواع داده‌ها

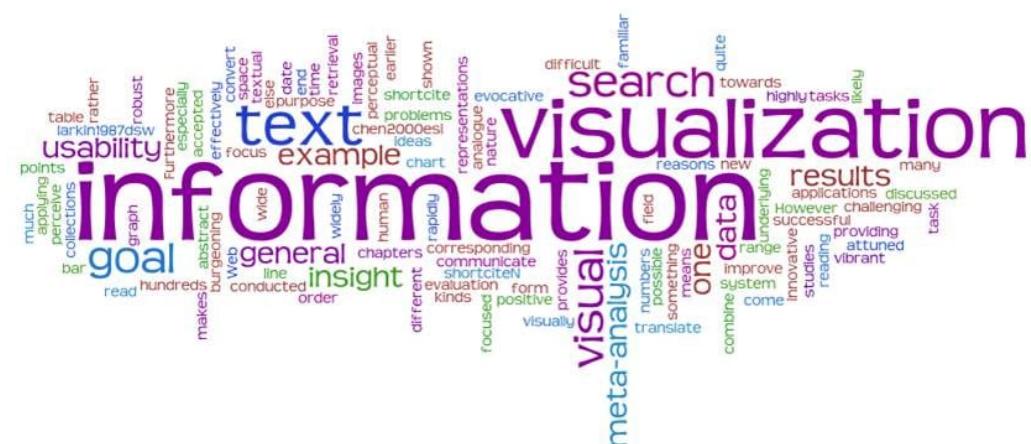
### داده‌های ساختاریافته Structured Data

### داده‌های غیرساختاریافته Unstructured Data

ویژگی‌های داده‌های ساختاریافته:

- سازمان‌یافته و منظم: داده‌ها دارای فرمت خاصی مانند جدول هستند
- سازگاری با پایگاه داده‌های رابطه‌ای: امکان ذخیره‌سازی و مدیریت آسان در پایگاه‌های داده سنتی
- بهینه برای جستجو و تحلیل: داده‌های ساختاریافته به سادگی قابل جستجو و پردازش هستند

## داده‌های غیرساختار یافته Unstructured Data



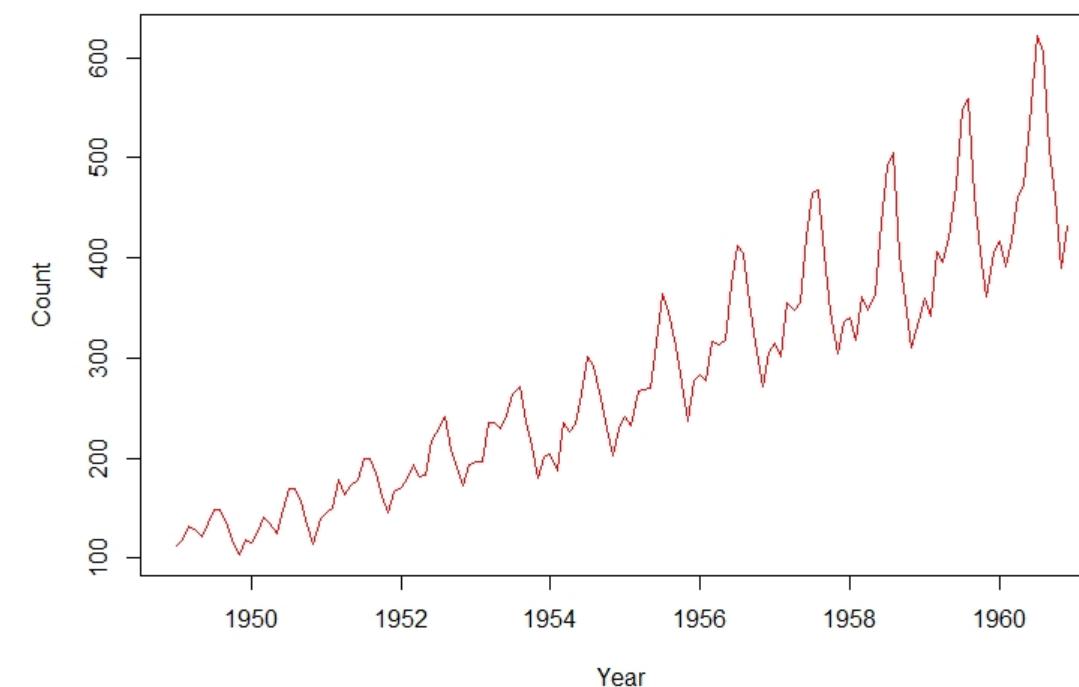
## داده‌های ساختار یافته Structured Data

## Tabular Data

ID	TOTAL ACTIONS	ACTION 1	ACTION 2	TOTAL TIME
10	120	80	40	0:50:05
11	255	130	125	1:40:03
12	180	100	80	1:20:19
13	305	205	100	1:58:58
14	71	50	21	0:35:41
15	418	310	108	2:08:18
16	222	150	72	1:32:58

# Time series data

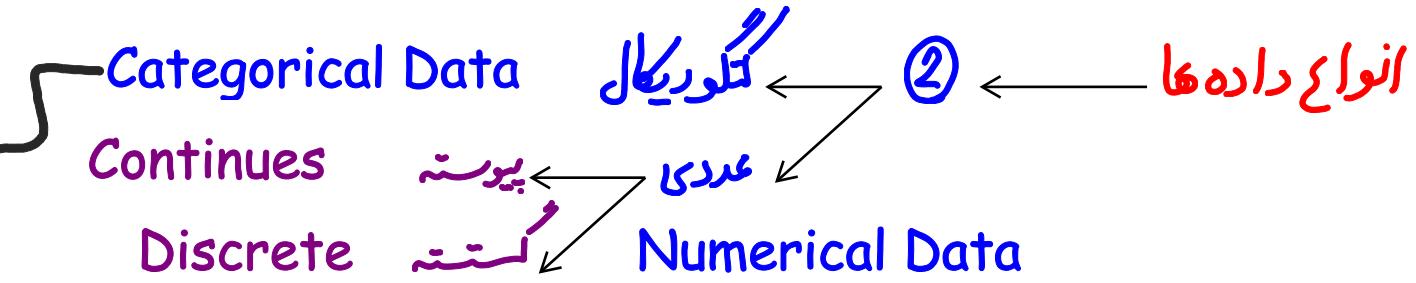
## Monthly Airline Passenger Numbers



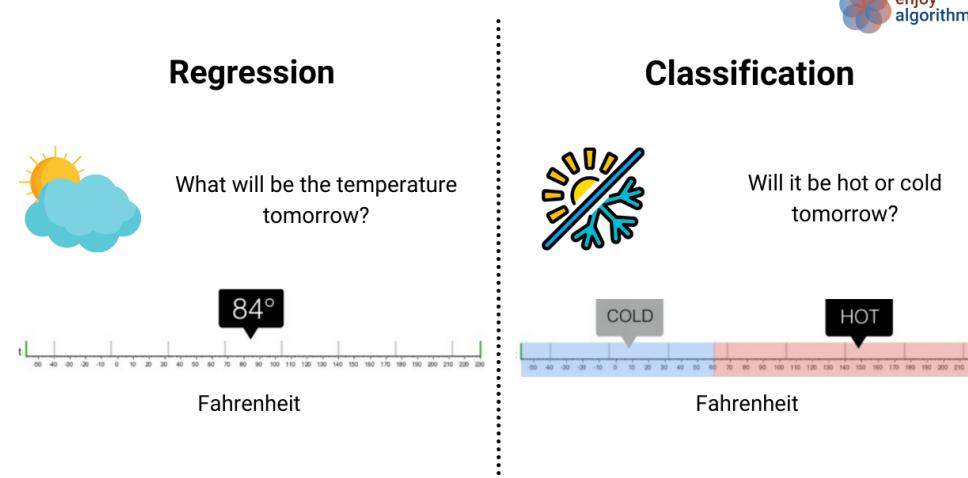
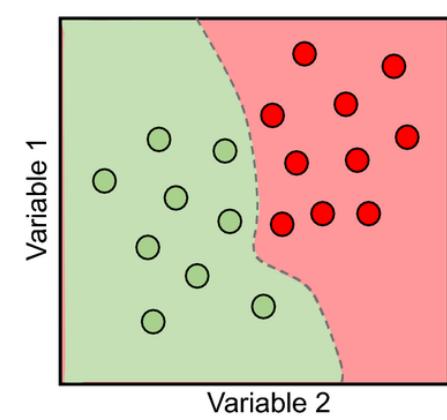
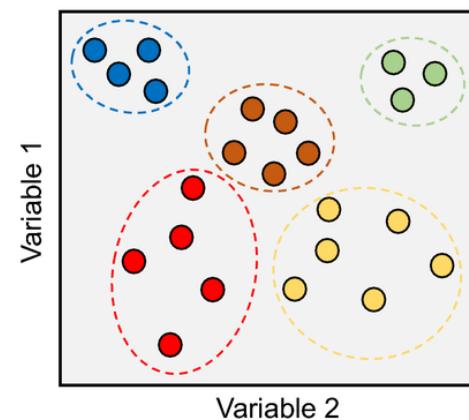
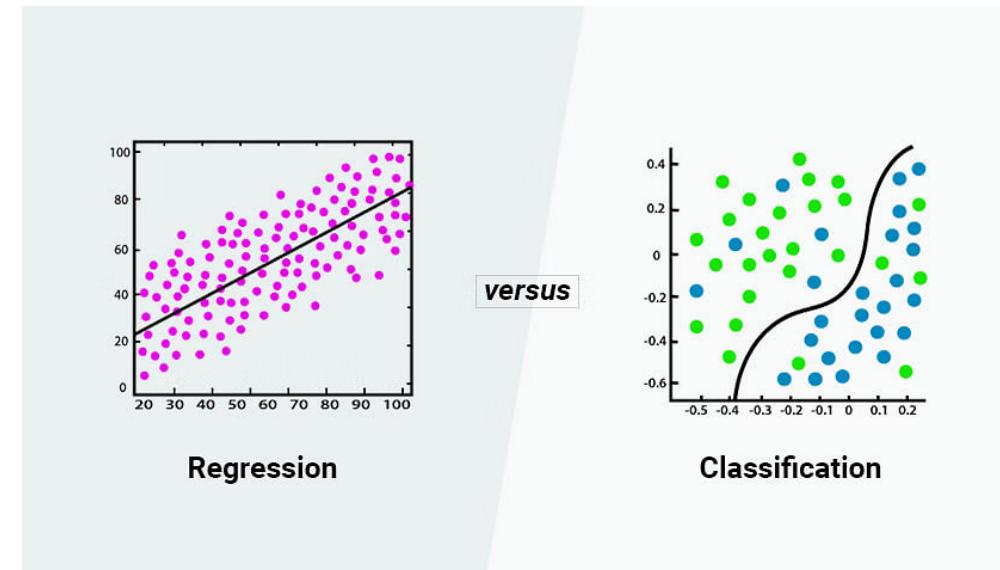
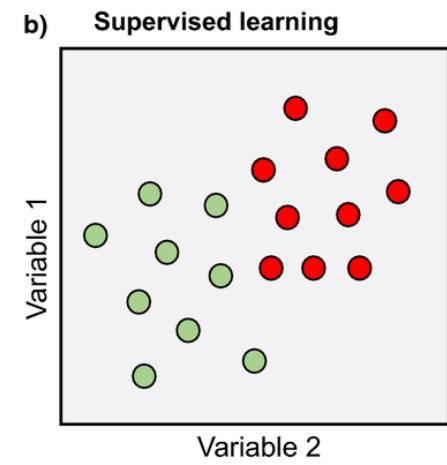
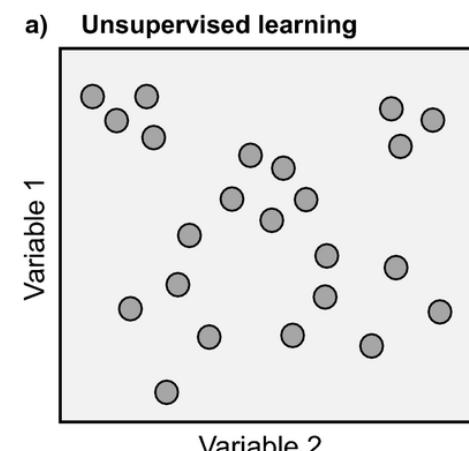
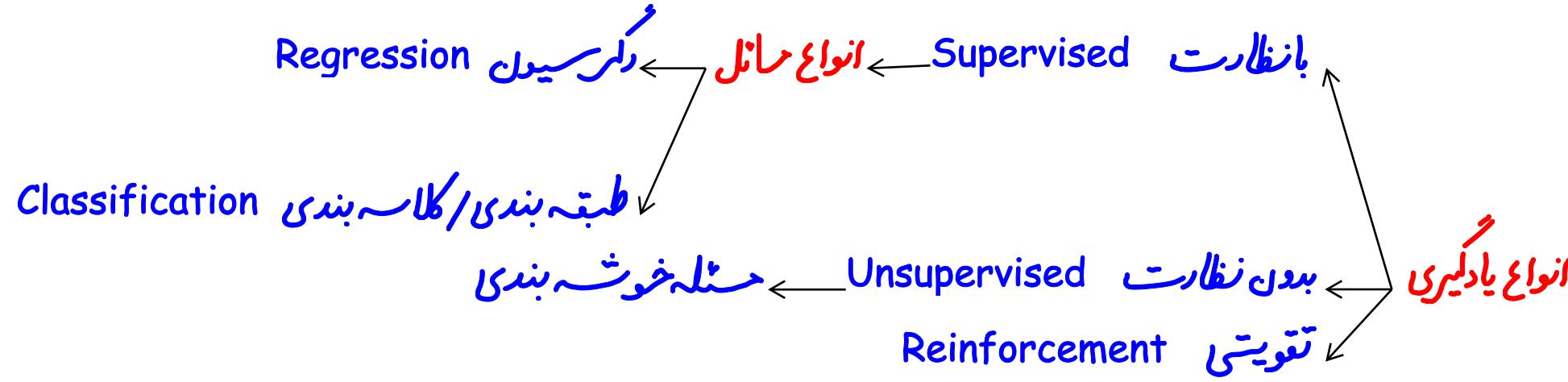
ویژگی‌های داده‌های غیرساختاری‌افتة  فاقد ساختار و قالب منظم: داده‌ها ممکن است شامل متون آزاد، فایل‌های چندرسانه‌ای یا ترکیبی از آن‌ها باشند نیاز به پردازش پیشرفتی برای تحلیل: تحلیل این داده‌ها به تکنیک‌های پردازش زبان طبیعی یا پردازش تصویر نیاز دارد سازگاری با پایگاه‌های داده: داده‌های غیرساختاری‌افتة معمولاً در پایگاه‌های داده غیررابطه‌ای ذخیره می‌شوند

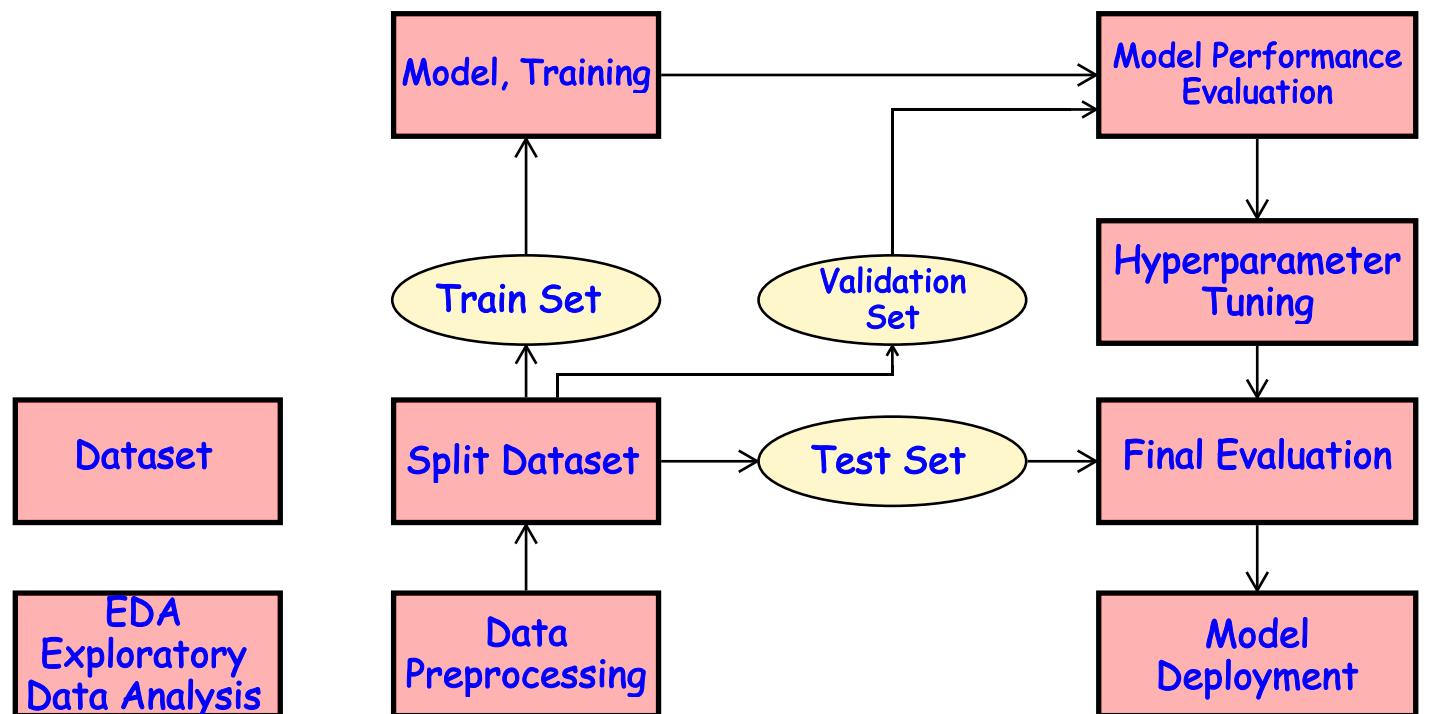
اسمی (Nominal) بدون ترتیب خاص، مانند رنگ‌ها

ترتیبی (Ordinal) دارای ترتیب مشخص، مانند سطوح تحصیلی (دیپلم، لیسانس)

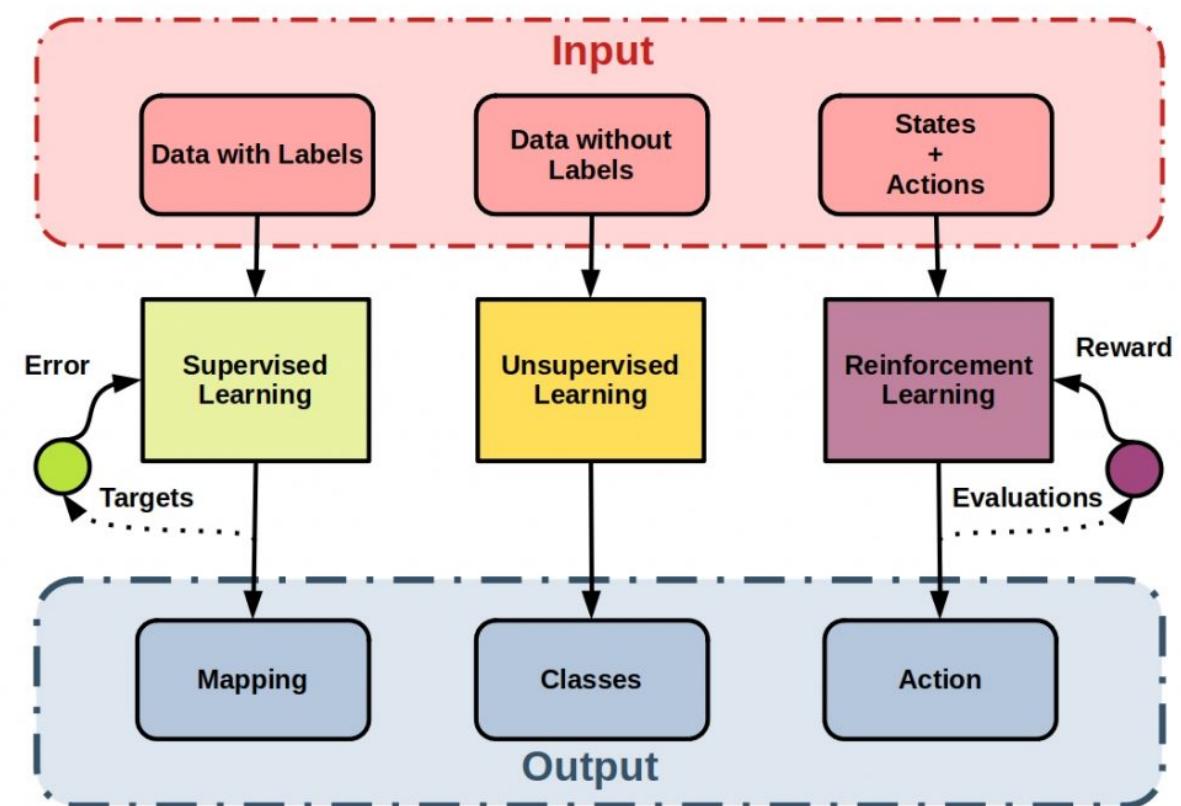


نوع داده	توضیحات	مثال‌ها	ویژگی‌ها
دسته‌ای (Categorical)	داده‌هایی که نشان‌دهنده دسته‌ها یا گروه‌ها هستند. این داده‌ها دارای ترتیب خاصی نیستند و نمی‌توانند مقادیر میانی یا عددی داشته باشند.	جنسیت (مرد، زن)، رنگ آبی، قرمز، سبز)، نوع خودرو (SUV، سدان)	مقادیر مجزا و غیرقابل اندازه‌گیری عددی
عددی (Numerical)	داده‌هایی که دارای مقادیر عددی هستند و می‌توان آن‌ها را اندازه‌گیری کرد یا روی آن‌ها عملیات ریاضی انجام داد.	قد، وزن، درآمد، تعداد اعضای خانواده	دارای مقادیر عددی و قابل اندازه‌گیری
- پیوسته (Continuous)	نوعی از داده‌های عددی که می‌تواند هر مقدار را در یک بازه مشخص بپذیرد و شامل اعداد اعشاری نیز می‌شود.	قد (۱۷۰.۰ سانتیمتر)، وزن (۳۶.۶ کیلوگرم)، دما (۴۰.۳ درجه)	بین‌نهایت مقادیر ممکن بین دو عدد مشخص
- گسته (Discrete)	نوعی از داده‌های عددی که فقط مقادیر مشخص و جدا از هم را می‌پذیرد و معمولاً شامل اعداد صحیح است.	تعداد فرزندان (۰، ۱، ۲، ...)، تعداد خودروها در خانواده	محدود به اعداد صحیح و شمارش‌پذیر





جمع‌آوری و آماده‌سازی داده‌ها: با استفاده از داده‌های خام، تحلیل اکتشافی انجام شده و داده‌ها پیش‌پردازش می‌شوند  
 تقسیم داده‌ها: مجموعه داده به بخش‌های آموزشی، اعتبارسنجی و آزمایشی تقسیم می‌شود  
 آموزش مدل: مدل روی داده‌های آموزشی آموزش داده می‌شود  
 ارزیابی و بهینه‌سازی مدل: عملکرد مدل با استفاده از داده‌های اعتبارسنجی ارزیابی و بهینه‌سازی می‌شود  
 ارزیابی نهایی و پیاده‌سازی: مدل روی داده‌های آزمایشی ارزیابی نهایی شده و سپس در دنیای واقعی پیاده‌سازی می‌شود



یادگیری ناظارت شده: در این نوع یادگیری، از داده‌های برچسب‌دار استفاده می‌شود. مدل با داده‌های ورودی و خروجی مشخص آموزش داده می‌شود و هدف آن یادگیری ارتباط بین داده‌ها است. خطا و هدف به مدل کمک می‌کنند تا عملکرد خود را بهبود دهد

یادگیری بدون ناظارت: در این نوع، داده‌ها بدون برچسب هستند و مدل تلاش می‌کند تا الگوهای کلاس‌ها را در داده‌ها شناسایی کند. هدف یادگیری ساختار یا الگوهای پنهان در داده‌ها بدون داشتن خروجی‌های مشخص است

یادگیری تقویتی: در این نوع، مدل از طریق تعامل با محیط و با استفاده از حالت‌ها و اقدامات آموزش می‌بیند. مدل با دریافت پاداش یا بازخورد یاد می‌گیرد که بهترین تصمیم‌ها یا اقدامات را انجام دهد

