

Date:

Sub:

بہ نام خدا

درس: مباحث ہیروہ

استاد: محمد احمد زاہد

مدرسہ اعلیٰ، مدرسہ انجیری - عالیہ مدرسہ اسلامیہ

A: خیر/ Cleaning Det + 2 در علم داده اهمیت دارد؟

با کمبودی داده‌ها یک مرحله ضروری در فرآیند علم داده است که به بهبود

کیفیت داده‌ها - افزایش دقت مدل‌ها - تصمیم‌گیری بهتر - صرفه جویی

در زمان و هزینه و جلوگیری از نتایج همراه کننده کند می‌شود

1. بهبود کیفیت داده‌ها 2. افزایش دقت مدل‌های یادگیری ماشین

3. تصمیم‌گیری بهتر 4. صرفه جویی در زمان و هزینه

5. جلوگیری از نتایج همراه کننده

Date:

Sub:

## B: Missing values چگونه مدیریت می شود؟

در علم داده یکی از مراحل مهم پیش پردازش داده ها است که تاثیر زیادی بر

کیفیت تحلیل و مدل سازی دارد. روش های مختلفی برای مدیریت این مقادیر

وجود دارد که بسته به نوع داده و مسئله انتخاب می شوند در ادامه چند راه

مقدار اول برای مدیریت مقادیر گم شده آورده شده است.

1 حذف داده های دارای مقادیر گم شده 2 جایگزینی مقادیر گم شده

3 علامت گذاری مقادیر گم شده 4 استفاده از روش های پیشرفته تر.

c: outliers چیست و چگونه می توانیم آن ها را تشخیص دهیم؟

outliers یا داده های بیرون نقطه داده ای هستند که به دلیل قابل توجهی

با سایر داده ها در مجموعه داده تفاوت دارند. این نقاط می توانند ناشی

از خطا در اندازه گیری، تغییرات غیر عادی در داده ها یا حتی نشان دهنده

پدیده های واقعی و مهم باشند.

روش های تشخیص outliers:

۱. روش های بصری ۲. روش های آماری ۳. روش های مبتنی بر فاصله

۴. روش های مدل سازی

Dexte Trans Formation :D

Data Types Formation به هر داده‌ای که در فرآیند علم داده است

کہ بہ بصیرت کیفیت راہہ ما اعتدائی کا اسی تحلیل بصیرت کہلہ ملے گا

یادگیری خاصین و انطباق با الزامات سیستم حکم می کند.

1. بصورت دقت داده 2. افزایش کارایی تحلیل 3. بهبود عملکرد مدل های یادگیری

ما نحن ؟ اطلبنا يا الزمات سيستم



## Encoding/Decoding Techniques (one-encoding: E

اظهار Label چه تفاوتی دارند؟

در پردازشی داده‌ها به خصوص در یادگیری ماشینی **Encoding** تکنیک‌های مختلفی

برای تبدیل داده‌های غیر عددی (مانند داده‌های دسته‌بندی یا متنی) به داده‌های

عددی استفاده می‌شوند.

**Label Encoding** یک تکنیک است که در آن هر دسته از داده‌ها به یک عدد

صحیح منحصربه‌فرد تبدیل می‌شود به عنوان مثال اگر یک ویژگی دسته‌بندی

به نام "رنگ" شامل مقادیر "قرمز"، "سبز" و "آبی" باشد، این مقادیر به شکل زیر

تبدیل می‌شود. **ارنگ**

Date:

Sub:

Feature Selection در Model-building اهمیت دارد؟

Feature Selection (انتخاب ویژگی) در فرآیند Model Building

(مدل سازی) اهمیت بسیار زیادی در Feature Selection به معنای انتخاب

زیر مجموعه ای از ویژگی های موجود در داده ها است که بیشترین تأثیر را بر عملکرد

مدل دارند و می توانند به ساخت یک مدل دقیق تر - کارآمدتر و قابل تفسیرتر

1. بهبود دقت مدل 2. افزایش کارایی مدل 3. بهبود قابلیت تفسیر مدل

4. کاهش اثر پیره 5. کاهش هزینه ها

Date:

Sub:

6: Duplicate Det چگونه در پایگاه داده ها حذف می شود؟

حذف داده های تکراری (Duplicate Det) در پایگاه های داده های مبتنی بر

بر اساس اطمینان از کیفیت و صحت داده ها است و وجود داده های تکراری

می تواند منجر به نتایج نادرست در تحلیل ها، گزارش ها و تصمیم گیری

ها شود.

1) شناسایی داده های تکراری 2 حذف داده های تکراری 3 جلوگیری از ایجاد داده های

تکراری





Date:

Sub:

Machine Learning Date: 11/11/1402  
موضوع: یادگیری ماشین

ایجاد مدل

داده‌های نامرتبط (Irrelevant Data) می‌توانند مشکلات

محدودری در پیش‌بینی‌های یادگیری ماشین ایجاد کنند. این داده‌ها شامل

ویژگی‌های هستند که هیچ ارتباط معناداری با متغیر هدف

(Target Variable) ندارند یا تأثیر بسیار کمی بر آن دارند.

1. کاهش دقت مدل 2. افزایش زمان آموزش و پیچیدگی

محاسباتی 3. کاهش قابلیت تفسیر مدل 4. افزایش ایجاد داده‌ها

Date:

Sub:

I: چه Data Imputation برای پر کردن missing values

کاربرد داده

Data imputation (جایگذاری داده) تکنیکی است که برای پر کردن مقادیر

گمشده (missing values) در مجموعه داده‌ها استفاده می‌شود.

(قطع داده) اصطلاحی است که در این زمینه به معنای ایجاد مقادیر آزمایشی

یا ارزیابی عملکرد روش‌های جایگذاری داده به کار می‌رود. نکته حوا

که روش جایگذاری باید



Date:

Sub:

ن: چگونه می توانیم  $NOx$  را در داده های عددی بررسی کنیم

بررسی نمرات یون (NOx) داده های عددی یک نام که در تحلیل داده ها

است زیرا بسیاری از روش های آماری (مانند رگرسیون  $NOx$  و

رگرسیون خطی) بر فرض نمرات یون داده های استوار هستند اگر داده ها

نمرات نباشند ممکن است نتایج این آزمون ها غیر قابل اعتماد باشند

1 روش های بررسی 2 روش های عددی