

1) কেন Date এবং Time হ্যান্ডেল করা জরুরি

একটা date column দেখতে শধু একটা তারিখ মনে হলেও, আসলে এর ভিতরে অনেক hidden signal থাকে। মডেল সরাসরি 2023-10-05 টাইপ তারিখ “বুবাতে” পারে না, কিন্তু তুমি এই তারিখ থেকে বের করা feature গুলো বুবাতে পারে।

Date থেকে সাধারণত যেসব তথ্য বের করা হয়

1. Year, Month, Day
কোন বছরে, কোন মাসে, মাসের কত তারিখে ঘটেছে
2. Day of week
Monday, Tuesday ইত্যাদি
ব্যবহারকারী আচরণ অনেক সময় weekday অনুযায়ী বদলায়
3. Weekend নাকি weekday
Saturday বা Sunday হলে আচরণ আলাদা হতে পারে
যেমন weekend এ spending বেশি
4. Quarter বা semester
Business এবং finance ডেটায় seasonality বোঝার জন্য কাজে লাগে
Q1 থেকে Q4, তারপর semester 1 বা 2

তুমি এখানে মূলত date কে ভেঙে ছোট ছোট meaningful parts বানিয়ে দিচ্ছো, যাতে মডেল pattern ধরতে পারে।

2) প্রথম কাজ Object থেকে Datetime এ কনভার্ট করা

Pandas এ dataset ইমপোর্ট করলে date column অনেক সময় string বা object টাইপ হয়ে যায়। তখন dt accessor কাজ করে না, এবং date-based calculation করা যায় না।

সমাধান

pd.to_datetime দিয়ে প্রি কলামকে datetime টাইপে convert করতে হয়।

কেন এটা দরকার

Datetime না হলে year, month, day বের করা যাবে না

Difference বা time delta বের করা যাবে না

Sorting, filtering, resampling সব কঠিন হবে

প্র্যাক্টিক্যাল নেট

কখনো date format inconsistent থাকলে parsing error হতে পারে

তখন errors, format, dayfirst এসব parameter লাগতে পারে

3) Date extraction কীভাবে করা হয় এবং কেন করা হয়

ধরা যাক Orders dataset এ একটি OrderDate কলাম আছে। তুমি এতে থেকে একাধিক feature বানাবে।

Year এবং Month

Year feature দিয়ে long-term trend ধরা যায়

Month feature দিয়ে seasonality ধরা যায়

Month number ব্যবহার করা যায়, আবার month name দিয়েও analysis সহজ হয়

Day of month এবং Day name

Day of month কিছু ক্ষেত্রে payroll, billing cycle, salary day এসব বোঝায়

Day name দিয়ে weekday pattern বোঝা যায়

Weekend flag

একটা boolean feature বানানো হয়

Weekend হলে 1, না হলে 0

এটা model কে সরাসরি একটা শক্ত signal দেয়

Quarter এবং Semester

Quarter দিয়ে yearly cycle কে 4 ভাগে দেখা যায়

Semester দিয়ে 2 ভাগে দেখা যায়

Business রিপোর্টিং বা demand forecasting এ এগুলো খুব common

ক্লাসিক ধারণা

একটা date column থেকে তুমি একসাথে trend, seasonality, weekly cycle, holiday behavior এসব signal তুলে আনতে পারো।

4) Time extraction কীভাবে করা হয় এবং কেন করা হয়

Messages dataset এ ধরে নাও timestamp আছে, যেমন 2026-01-29 19:45:20

এখান থেকে তুমি বের করতে পারো

Hour

কোন সময় message বেশি আসে

Minute

কিছু সিস্টেমে batching বা peak minute pattern থাকে

Second

High-frequency logs বা event stream এ কাজ করে

Only time

যদি date বাদ দিয়ে daily routine pattern ধরতে চাও

Time feature এর গুরুত্ব কোথায় বেশি

User activity analysis

Call center বা support load forecasting

Delivery বা ride demand prediction
System logs anomaly detection

5) Time difference বা elapsed time বের করা

এটা খুব গুরুত্বপূর্ণ কারণ অনেক prediction সমস্যাই recency-এর উপর নির্ভর করে।

উদাহরণ

আজকের তারিখ থেকে order date বাদ দিলে পাওয়া যায়

How many days since last order

এটা churn prediction, repeat purchase prediction, customer segmentation এ কাজে লাগে

Time delta থেকে কী কী বের করা যায়

Days elapsed

Hours elapsed

Seconds elapsed

Months elapsed

কিন্তু এখানে একটা সতর্কতা

Months কে days এর মতো fixed unit ধরা যায় না কারণ মাসের দিন সংখ্যা সমান নয়

তাই months বের করতে special handling লাগে

numpy timedelta বা period-based calculation ব্যবহার করা হয়

প্র্যাক্টিক্যাল টিপস

Recency feature বানানোর সময় reference date ঠিক করতে হয়

Training data এর ক্ষেত্রে future leakage যেন না হয়

অর্থাৎ prediction date এর পরের তথ্য ব্যবহার করা যাবে না

6) ভালো Date Time feature engineering করার কিছু নিয়ম

1. Timezone ঠিক আছে কি না
UTC বনাম local time mismatch হলে hour-based feature ভুল হয়ে যাবে
2. Missing বা invalid date কীভাবে handle হবে
NaT হলে fill বা drop strategy ঠিক করতে হবে
3. Cyclic encoding দরকার কি না
Month বা day of week আসলে cycle
12 এর পর আবার 1 আসে
সেক্ষেত্রে sin-cos encoding অনেক ক্ষেত্রে ভালো কাজ করে, বিশেষ করে linear model এ
4. Train-test leakage এ সাবধান
Today থেকে difference বের করলে “today” কোনটা তা consistent হতে হবে

সারসংক্ষেপ

Date time কে মডেলে ঢোকানোর সেরা কৌশল হলো

প্রথমে column কে datetime করা

তারপর date এবং time কে ভেঙে year, month, weekday, weekend, quarter, hour ইত্যাদি feature বানানো

এবং প্রয়োজনে recency বা elapsed time feature তৈরি করা