Pichle task mein humne features select kiye jo humne apne model ko train karne ke liye istemal kiye the. Ab hum sab categorical variables ko encode kar rahe hain aur train aur test set ko update kar rahe hain in encoded variables ke saath.

Is task ke ant mein, aapka preprocessing data complete ho jayega taki hume ek behtar performance mil sake jab hum model ko evaluate karenge aur usse build aur train karne ke liye tayar ho.

Jab aap apne model ko evaluate karte hain aur hum model ko build aur train karne ke liye tayar hote hain, toh categorical data ko transform karne ki zarurat ko samajhna zaroori hai. Kuch machine learning models, jaise ki support vector machine ya linear regression, categorical data ko input ke roop mein nahi lete hain.

Is limitation ke karan, hume categorical variables ko numerical values mein convert karna padta hai. Agar aap abhi tak samajh nahi paaye hain ki encoding categorical data kya hota hai, toh ek udaharan hai PClass humari feature. Yeh numbers 1, 2 aur 3 ko represent karte hain jo ki teen categories ko darshate hain. Aapki feature kabhi bhi in teeno values mein se zyada assume nahi karegi. Lekin kyunki yeh numbers hain, humare model ko inhe numerical data ke roop mein samajh aa jaata hai, jaise ki classes ka size, ek class, do classes ya teen classes. Aur agar yeh quantities hote aur types nahi, toh jab hum encoding karte hain, hum is category ke liye ek unique value dete hain taki model galat tarah se na interpret kare. Hum Get dummies ka istemal karenge categorical variables ko dummy indicator variables mein convert karne ke liye. Chaliye shuru karte hain feature Embarked, Sex, aur PClass ke saath.

Sabse pehle hum Embarked feature ko use karenge humare trained dataset mein aur hum prefix istemal karenge "embarked". Ab hum sex aur pclass ke saath bhi same process repeat karenge. Ise pause karein aur video ke instructions ke mutabiq embark feature ke saath wahi karein jo maine bataya. Other features ke saath bhi wahi process karein. Jaise hi humne encode kiya categorical features ko, unki nayi shape aisi hogi. Ab hum code ko run karenge aur agli steps mein humein ye naye encoded features ko training dataset mein add karna hoga. Concat ka istemal karte hain iske liye aur code ko run karte hain.

Jaise hi aap dekhte hain, hamare categorical features naye features ban gaye hain yahaan. Aap dekhte hain, humare paas saare categories naye features ke roop mein hain aur hum original features ko drop kar sakte hain kyunki ab yeh encode ho chuke hain. Drop karne ke baad woh aise dikhenge. Ab video ko pause karein aur apne training dataset mein jo kuch bhi kiya hai wahi apne test dataset mein bhi karein. Chaliye hum test dataset ke results dekhte hain. Hum code ko run karenge aur original features ko drop karenge. Ab humne data ka preprocessing complete kar liya hai. Humne missing data ka dhyaan rakha hai aur categorical features ko encode kiya hai. Agli task mein, humein trains aur test mein split karna hai aur humare model ko build karna hai. Agla task mein milte hain.