

Pada percobaan discretization dapat dilihat bahwa nilai MSE menurun.

```
Model dengan Fitur Kontinu:  
MSE: 24.2911  
R2: 0.6688  
  
Model dengan Fitur Discretisasi:  
MSE: 17.0118  
R2: 0.7680
```

Hal ini menunjukkan bahwa discretization berhasil menangkap pola atau hubungan dalam data yang tidak terlihat oleh model ketika menggunakan fitur kontinu. Berikut adalah analisis mendalam mengapa hal ini bisa terjadi:

1. Menangkap Hubungan Non-Linear

Penyebab: Fitur numerik kontinu mungkin memiliki hubungan non-linear dengan target. Model regresi linear (seperti LinearRegression) hanya dapat menangkap hubungan linear. Dengan melakukan discretization, fitur numerik diubah menjadi kategori yang dapat menangkap pola non-linear.

Contoh: Misalnya, fitur *rm* (jumlah kamar) mungkin memiliki hubungan non-linear dengan *medv* (harga rumah). Discretization membagi *rm* ke dalam beberapa interval (misalnya, 3-4 kamar, 5-6 kamar, dst.), sehingga model dapat mempelajari pola yang berbeda untuk setiap interval.

2. Mengurangi Noise dalam Data

Penyebab: Fitur numerik kontinu mungkin mengandung noise atau variasi kecil yang tidak relevan dengan target. Discretization dapat mengurangi noise ini dengan mengelompokkan nilai-nilai yang mirip ke dalam kategori yang sama.

Contoh: Jika fitur *lstat* (persentase populasi berstatus ekonomi rendah) memiliki variasi kecil yang tidak signifikan, discretization dapat mengelompokkan nilai-nilai ini ke dalam kategori yang lebih besar, sehingga model fokus pada perbedaan yang lebih signifikan.

3. Menangani Outlier

Penyebab: Fitur numerik kontinu mungkin mengandung outlier yang dapat memengaruhi performa model. Discretization dapat mengurangi dampak outlier dengan mengelompokkannya ke dalam kategori yang sesuai.

Contoh: Jika fitur age (usia properti) memiliki beberapa nilai yang sangat tinggi (outlier), discretization dapat mengelompokkan nilai-nilai ini ke dalam kategori "usia tua", sehingga outlier tidak memengaruhi model secara berlebihan.

4. Meningkatkan Interpretabilitas

Penyebab: Discretization mengubah fitur numerik menjadi kategori yang lebih mudah diinterpretasi oleh model. Ini dapat membantu model mempelajari hubungan yang lebih jelas antara fitur dan target.

Contoh: Fitur rm yang diubah menjadi kategori seperti "kamar sedikit", "kamar sedang", dan "kamar banyak" lebih mudah diinterpretasi oleh model daripada nilai numerik yang kontinu.

5. Efek Regularisasi Implisit

Penyebab: Discretization dapat berfungsi sebagai bentuk regularisasi dengan mengurangi kompleksitas model. Dengan mengurangi jumlah nilai unik dalam fitur, discretization mencegah model dari overfitting.

Contoh: Jika fitur lstat memiliki banyak nilai unik, discretization dapat mengelompokkannya ke dalam beberapa kategori, sehingga model tidak terlalu kompleks dan lebih generalizable.

6. Menangkap Interaksi antara Fitur

Penyebab: Discretization dapat membantu model menangkap interaksi antara fitur-fitur yang sulit dipelajari ketika fitur tersebut dalam bentuk kontinu.

Contoh: Interaksi antara rm (jumlah kamar) dan lstat (status ekonomi) mungkin lebih mudah dipelajari ketika kedua fitur diubah menjadi kategori.