

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan pembahasan hasil yang telah diuraikan, kesimpulan pada skripsi ini adalah:

1. Model pohon regresi pertama yang dibangun, tidak dibatasi oleh ukuran apapun. Model ini melibatkan seluruh variabel numerik, yaitu harga tanah per meter persegi, luas tanah, luas bangunan, tingkat rumah, banyak kamar mandi, banyak kamar tidur, dan daya listrik, dengan visualisasi yang dihasilkan sangat rumit, serta memiliki akurasi sebesar 80%. Dengan demikian, berdasarkan standar model yang baik yang telah diuraikan sebelumnya, model ini belum mencapai hal tersebut karena masih tergolong *overfitting*. Jadi, model pohon regresi ini belum dapat digunakan untuk memprediksi harga jual rumah di Kota Bandung dengan baik, sehingga perlu diperbaiki lagi.
2. Model pohon regresi kedua yang dibangun, dilakukan dengan mengevaluasi terlebih dahulu uji *underfitting* dan *overfitting*. Berdasarkan uji tersebut, model berada di kondisi yang baik jika banyak sampel minimum untuk pemecahan rekursif biner berkisar antara 14 sampai 17 atau 22 sampai 70. Dengan banyak sampel minimum untuk pemecahan rekursif biner sebesar 70, model pohon regresi yang dihasilkan hanya mengutamakan 3 aspek dari variabel numerik, yaitu luas bangunan, luas tanah, dan harga tanah per meter persegi. Dalam hal ini, model mengelompokkan berbagai keadaan menjadi 10 bagian, dengan nilai prediksi harga jual rumah yang ditunjukkan secara keseluruhan berkisar antara Rp 500.000.000,00 sampai Rp 18.000.000.000,00. Model ini memiliki akurasi sebesar 53%, di mana meskipun sudah tidak tergolong *overfitting*, akurasi model menjadi lebih rendah dari model pertama, sehingga masih belum dapat digunakan untuk memprediksi harga jual rumah di Kota Bandung dengan baik.
3. Model pohon regresi ketiga yang dibangun, dilakukan dengan mengkombinasikan model pohon regresi kedua dan regresi kernel, sehingga model disebut sebagai pohon regresi kernel. Model ini melibatkan seluruh variabel numerik seperti pada model pohon regresi pertama. Akan tetapi, model ini hanya diperuntukkan 6 bagian pada pohon regresi, sehingga untuk 4 bagian lainnya, proses prediksi langsung berdasarkan pada pohon regresi saja. Nilai prediksi harga jual rumah yang ditunjukkan model secara keseluruhan berkisar antara Rp 400.000.000,00 sampai Rp 13.500.000.000,00, dengan akurasi sebesar 78%. Dengan kondisi demikian, model yang sudah tidak tergolong *overfitting* juga menjadi memiliki akurasi lebih tinggi dari model kedua, sehingga sudah mencapai standar model yang baik dan dapat digunakan untuk memprediksi harga jual rumah di Kota Bandung dengan baik.

5.2 Saran

Berdasarkan isi pada skripsi ini, saran untuk pengembangan penelitian selanjutnya adalah:

1. Melakukan perhitungan untuk mengetahui variabel yang paling mempengaruhi pada model regresi kernel, sehingga variabel yang akan digunakan pada model keseluruhan menjadi lebih efektif.
2. Mengikutsertakan variabel kategorik ke dalam model regresi kernel, sehingga analisis yang dilakukan mengenai data tersebut tidak hanya berdasarkan statistika deskriptif.
3. Menggunakan fungsi kernel jenis lainnya, agar dapat diketahui perbandingan model regresi kernel yang lebih baik untuk dikombinasikan dengan pohon regresi.

DAFTAR REFERENSI

- [1] Soetedjo, E. M. (2020) *The King of Property 36 Secrets of Powerful Closing*. Elex Media Komputindo, Indonesia.
- [2] Dagar, A. dan Kapoor, S. (2020) *A Comparative Study on House Price Prediction*. *International Journal for Modern Trends in Science and Technology*, **6**.
- [3] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An Introduction to Statistical Learning with Application in R*. Springer, USA.
- [4] Torgo, L. (1999) *Kernel Regression Tree*. *Research Gate* (https://www.researchgate.net/publication/2378537_Kernel_Regression_Trees), **1**.
- [5] Motulsky, H. dan Christopoulos, A. (2003) *Fitting Models to Biological Data using Linear and Nonlinear Regression*, 4 edition. *GraphPad Prism*, USA.
- [6] Kumar, V. dan Garg, M. L. (2018) *Predictive Analytics: A Review of Trends and Techniques*. *International Journal of Computer Applications*, **182**.
- [7] Koerse, D. P., Botev, Z. I., Taimre, T., dan Vaisman, R. (2020) *Data Science and Machine Learning - Mathematical and Statistical Methods*. CRC Press, USA.
- [8] Abraham, B. dan Ledolter, J. (2006) *Introduction to Regression Modeling*. Brooks/Cole - Cengage Learning, USA.
- [9] Heryana, A. (2020) Uji statistik nonparametrik. *Research Gate* (https://www.researchgate.net/publication/341539787_UJI_STATISTIK_NON_PARAMETRIK), **1**.
- [10] Sarkar, D., Bali, R., dan Sharma, T. (2018) *Practical Machine Learning with Python A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Apress, India.
- [11] Han, J., Kamber, M., dan Pei, J. (2012) *Data Mining Concepts and Techniques*, 3 edition. Morgan Kaufmann Publishers, USA.
- [12] Carter, N. (2020) *Data Science for Mathematicians*, 1 edition. CRC Press, USA.
- [13] Ma, X. (2018) *Using Classification and Regression Trees A Practical Primer*. Information Age Publishing, Inc., USA.
- [14] Berk, R. A. (2016) *Statistical Learning from a Regression Perspective*, 2 edition. Springer, Switzerland.
- [15] W, H. S. D. K., Wilandari, Y., dan Sudarno (2012) Analisis faktor-faktor yang mempengaruhi indeks mutu benang menggunakan metode pohon regresi (studi kasus di pt. industri sandang nusantara unit patal grati). *Media Statistika*, **5**.
- [16] Gramacki, A. (2017) *Nonparametric Kernel Density Estimation and Its Computational Aspects*, 37 edition. Springer, Switzerland.

- [17] Silverman, B. W. (1998) *Density Estimation for Statistics and Data Analysis*. CRC Press, USA.
- [18] Schimek, M. G. (2000) *Smoothing and Regression - Approaches, Computation, and Its Application*. A Wiley-Interscience Publication, Canada.