

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dalam menganalisa himpunan data US Health Insurance menggunakan regresi linear, kuantil, polinomial, dan *spline*, diperoleh kesimpulan:

1. Model-model regresi linear, kuantil, polinomial, dan *spline* dapat digunakan untuk memodelkan besar klaim.
2. Dalam hasil regresi *spline*, karena variabel-variabel dibagi menjadi beberapa segmen, patut dicatat bahwa berdasarkan model tersebut, peningkatan usia dan BMI nasabah cenderung meningkatkan besar klaim. Secara keseluruhan, semua variabel dapat dikatakan cukup signifikan untuk diimplementasikan ke dalam model dengan status merokok sebagai variabel yang paling memengaruhi besar klaim nasabah. Di sisi lain, proses *stepwise* menghilangkan variabel *gender* dan *region* sehingga relatif kurang signifikan dibandingkan dengan variabel-variabel lain.
3. Berdasarkan hasil yang diperoleh, tanpa mempertimbangkan banyaknya parameter, regresi *spline* adalah model yang terbaik karena menghasilkan RMSE terkecil. Sebaliknya, banyak parameter menjadi pertimbangan, regresi linear dengan interaksi merupakan model terbaik. Selain itu, berdasarkan RMSE, kuantil yang paling cocok untuk regresi kuantil adalah kuantil ke-45. Terakhir, tetapi tidak kalah penting, untuk regresi polinomial, variabel *age* dan *children* memiliki hubungan kuadratik dan variabel *bmi* memiliki hubungan kubik terhadap *charges*.

5.2 Saran

Berikut beberapa pengembangan yang dapat dilakukan:

1. Dalam skripsi ini, interaksi hanya diimplementasikan untuk regresi linear. Hasil menunjukkan bahwa kedua hal ini meningkatkan performa model dan berpotensi untuk meningkatkan performa model lain. Ada baiknya apabila interaksi juga diimplementasikan ke dalam model-model agar diperoleh hasil yang terbaik.
2. Berdasarkan kesimpulan yang diperoleh, terjadi kebingungan dalam memilih model yang terbaik akibat kesimpulan yang berbeda antara ukuran evaluasi. Kedepannya, diharapkan dapat dicari solusi untuk permasalahan ini, seperti dengan menggunakan ukuran lain yang dapat memberikan kesimpulan yang lebih menyeluruh.
3. Meskipun dapat mencakup hubungan yang kuadratik, regresi-regresi yang digunakan tergolong linear. Penurunan model seperti regresi *penalized spline* juga berpotensi untuk memberikan hasil yang baik karena dengan menambahkan penalti, regresi *penalized spline* cenderung untuk tidak *overfit* dibandingkan dengan regresi *spline* biasa sehingga diluar menghasilkan model yang lebih baik, juga dapat mengurangi kemunculan kasus *overfit* [19]. Selain regresi *penalized spline*, juga dapat dipertimbangkan regresi nonparametrik seperti regresi lokal (LOESS), Generalized Additive Model (GAM), atau model-model nonparametrik lainnya.
4. Berbagai metode seperti analisis deret waktu, *decision tree* dan *random forest*, serta metode-metode lainnya tersedia. Hasil pengujian yang diperoleh dengan menggunakan regresi seperti pada skripsi ini ada baiknya dibandingkan dengan metode lain agar diperoleh hasil terbaik berdasarkan metode-metode yang digunakan.

DAFTAR REFERENSI

- [1] Smolka, A. (2003) The principle of risk partnership and the role of insurance in risk. *Disaster Risk Reduction in South Asia*, **13**, 37.
- [2] Galeotti, M., Gürtler, M., dan Winkelvos, C. (2013) Accuracy of premium calculation models for cat bonds—an empirical analysis. *Journal of Risk and Insurance*, **80**, 401–421.
- [3] Kwak, S. K. dan Kim, J. H. (2017) Statistical data preparation: Management of missing values and outliers. *Korean Journal of Anesthesiology*, **70**, 407–411.
- [4] Sukestiyarno, Y. dan Agoestanto, A. (2017) Batasan prasyarat uji normalitas dan uji homogenitas pada model regresi linear. *Unnes Journal of Mathematics*, **6**, 168–177.
- [5] Staffa, S. J., Kohane, D. S., dan Zurakowski, D. (2019) Quantile regression and it's applications: A primer for anesthesiologists. *Anesthesia & Analgesia*, **128**, 820–830.
- [6] Wolny-Dominiak, A., Ornat-Acedaska, A., dan Trzpiot, G. (2012) Insurance portfolios rate making: Quantile regression approach. *Proceedings of 30th International Conference Mathematical Methods in Economics, Karviná, Czech Republic*.
- [7] Stimson, J. A., Carmines, E. G., dan Zeller, R. A. (1978) Interpreting polynomial regression. *Sociological Methods & Research*, **6**, 515–524.
- [8] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An Introduction to Statistical Learning*. Springer.
- [9] (2014) Undang-undang (UU) tentang perasuransian. LN.2014/No. 337, TLN No. 5618, LL SETNEG: 60 HLM.
- [10] Guntara, D. (2016) Asuransi dan ketentuan-ketentuan hukum yang mengaturnya. *Justisi Jurnal Ilmu Hukum*, **1**, 29–46.
- [11] Guo, J., Zhou, S., Chen, J., dan Chen, Q. (2021) How information technology capability and knowledge integration capability interact to affect business model design: A polynomial regression with response surface analysis. *Technological Forecasting and Social Change*, **170**, 120935.
- [12] Sykes, A. O. (1993) An introduction to regression analysis. *Coase-Sandor Institute for Law & Economics Working Paper*, **20**.
- [13] Frees, E. W. (2009) *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press.
- [14] Koenker, R. (2017) Quantile regression: 40 years on. *Annual Review of Economics*, **9**, 155–176.
- [15] Irawan, R. (2020) Varying coefficient model for dengue fevers using p-splines quantile regression. Thesis. Hasselt University, Belgium.

- [16] Chiandussi, G., Bugeda, G., dan Oñate, E. (2000) Shape variable definition with c_0 , c_1 and c_2 continuity functions. *Computer methods in applied mechanics and engineering*, **188**, 727–742.
- [17] Marsh, L. C. (1983) On estimating spline regressions. *Proceedings of SAS Users Group International*, **8**, 723–728.
- [18] Chai, T. dan Draxler, R. R. (2014) Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific model development*, **7**, 1247–1250.
- [19] Brezger, A. dan Steiner, W. J. (2008) Monotonic regression based on bayesian p-splines: An application to estimating price response functions from store-level scanner data. *Journal of business & economic statistics*, **26**, 90–104.

