

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan pada bab-bab sebelumnya, diperoleh beberapa kesimpulan sebagai berikut.

1. Faktor-faktor (variabel) yang mempengaruhi data tingkat pemulihan kehamilan dilakukan dengan menggunakan *variable importance*. Variabel-variabel yang mempengaruhi, yaitu *Duration*, *Age Band*, *Integration with STD*, *Gross Indexed Benefit Amount*, dan *Own Occupation To Any Transition*.
2. Pada metode *Regression Trees* dan *Random Forests* memiliki variabel terpenting (variabel yang paling berpengaruh pada pohon) yang sama, yaitu waktu (*Duration*) dan umur (*Age Band*).
3. Metode *Random Forests* menghasilkan prediksi tingkat pemulihan kehamilan dengan melakukan *Bootstrap* dan *Out-of-Bag Error Estimation* serta mencari nilai prediksi (MSE) dari variabel-variabel yang digunakan.
4. Metode *Random Forests* menghasilkan prediksi data GLTD *Recovery Rates* lebih baik daripada metode *Regression Trees*. Hal ini dikarenakan memiliki jumlah variabel prediktor yang memadai dan jumlah data yang cukup besar untuk membuat beberapa simulasi *tree* dengan cara *Bootstrapping*, sedangkan metode *Regression Trees* hanya mengkategorikan satu *tree* dari keseluruhan data. Dengan demikian, hasil prediksi tersebut dapat dilihat dari *Mean Square Error* yang telah dihitung dan diperoleh nilai *Mean Square Error* dari penggunaan metode *Random Forests* lebih kecil daripada metode *Regression Trees*.

5.2 Saran

Berikut adalah beberapa saran pengembangan topik untuk penelitian selanjutnya.

1. Menggunakan sampel data dengan jumlah observasi dan variabel yang digunakan seimbang dan dilihat apakah menghasilkan nilai prediksi yang lebih baik.
2. Menggunakan algoritma *Machine Learning* yang lain, seperti metode *Support Vector Machines* (SVM), *k – Nearest Neighbors* (KNN), dan *Naive Bayes* untuk dibandingkan performansinya dalam mengestimasi nilai dari variabel yang digunakan dan menganalisis hasil yang diperoleh dengan menghitung *Mean Square Error* (MSE) yang dihasilkan untuk setiap metode.

DAFTAR REFERENSI

- [1] Zain, I. dan Akbar, Y. R. (2020) Asuransi. *Bank Dan Lembaga Keuangan Lainnya*. Deepublish.
- [2] Haberman, S. dan Pitacco, E. (2018) *Actuarial Models for Disability Insurance*. Routledge.
- [3] Kopinsky, M. (2017). *Predicting Group Long Term Disability Recovery and Mortality Rates using Tree Models*, Schaumburg: **The Society of Actuaries**.
- [4] Herbrich, R. dan Graepel, T. (2017) *Introduction to Machine Learning with Applications in Information Security*. CRC Press, California.
- [5] Nikolaev, N. dan Iba, H. (2008) *Adaptive Learning of Polynomial Networks: Genetic Programming, Backpropagation, and Bayesian Methods*. Springer, New York, USA.
- [6] Geron, A. (2017) *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 1st edition.
- [7] Nisbet, R., Miner, G., dan Yale, K. (2018) *Handbook of Statistical Analysis and Data Mining Applications*, 2nd edition.
- [8] Siyoto, S. dan Sodik, M. A. (2015) Komponen-komponen penelitian. Bagian dari Ayup (ed.), *Dasar Metodologi Penelitian*. Literasi Media Publishing.
- [9] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An Introduction to Statistical Learning with Applications in R*, . New York: Springer.
- [10] Gan, G. dan Valdez, E. A. (2018) Decision trees. *Actuarial Statistics with R: Theory and Case Studies*.
- [11] Hardle, W. dan Simar, L. (2007) *Applied Multivariate Statistical Analysis*, 2nd edition. Springer.
- [12] Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.
- [13] Suparman (2012) *Pengantar Bootstrap dan Aplikasinya*, 1st edition. JPMIPA FKIP UAD Press, Yogyakarta.
- [14] Efron, B. dan Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [15] Hastie, T., Tibshirani, R., dan Friedman, J. (2008) *The Elements of Statistical Learning*, 2nd edition. Springer, New York.

