

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berikut adalah beberapa kesimpulan yang diperoleh setelah dilakukan analisis dan perbandingan model *random forests* terhadap *decision tree*:

1. Berdasarkan model plot *random forests*, nilai galat cenderung stabil setelah *tree* ke-200. Dapat disimpulkan bahwa nilai parameter *ntree* dapat kurang dari 500.
2. Berdasarkan hasil *tuning random forests*, disimpulkan parameter *mtry* paling berpengaruh dalam hasil pencocokan model. Semakin besar nilai parameter *mtry* maka galat model diketahui akan semakin kecil dan diperoleh hasil pada model ini yang paling baik disaat *mtry* sama dengan jumlah variabel prediktor.
3. Model *random forests* secara keseluruhan lebih baik digunakan untuk menangani data dengan variabel prediktor yang banyak, dibandingkan dengan *decision tree*. Hal ini ditunjukkan pada tingkat *importance* yang relatif merata bagi setiap variabel prediktor.
4. Dari hasil model prediksi kematian peserta asuransi cacat diketahui bahwa, kategori cacat mempengaruhi model *random forest* sebesar 37%, durasi mempengaruhi model *random forest* sebesar 16%, umur mempengaruhi model *random forest* sebesar 15%, jenis kelamin mempengaruhi model *random forest* sebesar sebesar 10%, jenis pekerjaan mempengaruhi model *random forest* sebesar 9%, dan indeks gaji bulanan mempengaruhi model *random forest* sebesar 10%.
5. Model *random forests* tidak memberikan hasil prediksi tingkat mortalita GLTD 2017 yang lebih baik daripada *decision tree*. Hal ini dapat dilihat dari nilai *mean squared errors* yang telah dihitung, dan diperoleh nilai *mean squared errors* dari penggunaan model *decision tree* lebih kecil daripada model *random forests*.
6. Model *random forests* tidak lebih baik digunakan dalam data GLTD 2017 karena terdapat banyak sekali entri data dan kurang jumlah dari prediktornya.

6.2 Saran

Berikut adalah beberapa saran pengembangan topik untuk penelitian selanjutnya:

1. Prediksi tingkat mortalita GLTD 2017 menggunakan model lain seperti model *Bagging* atau model *Boosting*, lalu melakukan perbandingan *mean squared errors* (MSE) dengan hasil [5].
2. Prediksi tingkat mortalita menggunakan metode *random forest* dengan melakukan ekstraksi data lebih lanjut terhadap data GLTD 2017 dengan mengambil data dengan durasi tahunan serta mengkalkulasi ulang nilai *exposures*.

DAFTAR REFERENSI

- [1] Gan, G. dan Valdez, E. (2018) *Actuarial Statistics with R: Theory and Case Studies*, 1st edition. Actex Learning, Boston.
- [2] Martin, R. L. dan at all (2011) Experience studies - 2008 long-term disability experience table report. <https://www.soa.org/resources/experience-studies/2011/2008-ltd-experience-report/>. 15 Oktober 2019.
- [3] Mitchell, T. (1997) *Machine Learning*, 1st edition. McGraw-Hill Science, New York.
- [4] Horning, N. (2010) Introduction to decision trees and random forests. *American Museum of Natural History Center for Biodiversity and Conservation*, **1**, 1–6.
- [5] Kopinsky, M. (2017) Predicting group long term disability recovery and mortality rates using tree models. *Society of Actuaries*, **1**, 1–31.
- [6] Hill, D. B. (1987) Employer-sponsored long-term disability insurance. *Bureau of Labor Statistics*, **7**, 16–22.
- [7] Samuel, A. (1959) Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, **3**, 210–229.
- [8] Cutler, A., Curler, D. R., dan Stevens, J. R. (2011) Random forests. *Machine Learning*, **1**, 1–20.
- [9] Orloff, J. dan Bloom, J. (2014) Variance of discrete random variables. *MIT OpenCourseWare*, **5**, 1–7.
- [10] Breiman, L. (2001) Random forest. *Statistics Department, University of California*, **1**, 1–33.
- [11] Biau, G. dan Scornet, E. (2015) A random forest guided tour. *Sorbonne Universites*, **1**, 1–41.
- [12] Breiman, L., Cutler, A., Liaw, A., dan Wiener, M. (2018) Breiman and cutler’s random forests for classification and regression. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>. 22 Maret 2020.

