

## BAB V

### SIMPULAN DAN SARAN

#### 5.1 Simpulan

Perusahaan *P2P Lending* memerlukan model *machine learning* guna mengklasifikasikan intensitas penagihan nasabah dalam *Collection Intensity Scoring (CIS)* dan memprediksi jumlah interaksi yang akan direspon nasabah pada *communication channel* yang direkomendasikan. Data yang digunakan dalam penelitian ini merupakan data nasabah dari perusahaan *P2P Lending* yang diolah dengan bahasa pemrograman PostgreSQL (PL/pgSQL) melalui *software* DBEaver untuk kebutuhan *modeling* dalam Python. Penelitian ini menerapkan dua algoritma berdasarkan jenis pendekatannya, yaitu algoritma *Random Forest* untuk pendekatan klasifikasi terhadap intensitas penagihan nasabah dan algoritma *K-Nearest Neighbors Regressor* untuk pendekatan prediksi terhadap jumlah interaksi yang akan diperoleh nasabah pada *channel* yang direkomendasikan. Penerapan model *Random Forest* melibatkan data pribadi dan pinjaman nasabah, serta kategori intensitas penagihan berdasarkan *scoring* yang ditetapkan, yaitu *low*, *medium*, dan *high*. Penghitungan *intensity scoring* ditetapkan berdasarkan jumlah interaksi pada setiap *communication channel*, jumlah *loan* yang terbayarkan, dan rata-rata durasi dari waktu interaksi menuju pembayaran. Sementara itu, penerapan model *K-Nearest Neighbors* memberikan rekomendasi *communication channel* berdasarkan perilaku interaksi nasabah, yakni antara *email*, *robocall*, *SMS*, dan *telephony*. Hasil rekomendasi *channel* beserta jumlah interaksi tersebut kemudian diprediksi menggunakan *K-Nearest Neighbors Regressor*.

Penelitian ini menerapkan *hyperparameter optimization* dengan metode *Grid Search* dan *Bayesian Optimization* pada model *Random Forest* dan *K-Nearest Neighbors Regressor*. Penerapan *Grid Search* dan *Bayesian Optimization* dilakukan untuk meningkatkan kinerja dari kedua model konvensional. Melalui *Grid Search*, dilakukan eksplorasi secara ekstensif terhadap kombinasi *hyperparameter* di ruang pencarian yang ditentukan. *Hyperparameter* terbaik yang diperoleh berhasil ditemukan yang tidak hanya meningkatkan akurasi pada *Random*

*Forest*, tetapi juga mengurangi *error* pada prediksi *K-Nearest Neighbors Regressor* melalui *hyperparameter* terbaik, meskipun dengan waktu komputasi yang lebih lama. Di sisi lain, *Bayesian Optimization* memanfaatkan probabilitas model guna mendapatkan informasi dari hasil evaluasi sebelumnya untuk memungkinkan pencarian dan pemerolehan *hyperparameter* yang lebih optimal. Metode ini berhasil menemukan *hyperparameter* yang menghasilkan akurasi yang tinggi pada *Random Forest* dan *error* yang rendah pada *K-Nearest Neighbors Regressor* dalam waktu yang singkat.

Penelitian ini membandingkan kinerja model *Random Forest* dan *K-Nearest Neighbors Regressor*, baik dalam bentuk konvensional maupun setelah menerapkan *hyperparameter optimization* dengan metode *Grid Search* dan *Bayesian Optimization*. Perbandingan ini dilakukan dengan fokus pada 2 *evaluation metrics* utama, yaitu *accuracy* untuk *Random Forest* dan *Mean Absolute Error (MAE)* untuk *K-Nearest Neighbors Regressor*. Dalam pendekatan klasifikasi, model *Random Forest* menunjukkan nilai *accuracy* sebesar 98.16%. Setelah penerapan *Grid Search*, terjadi peningkatan nilai *accuracy* menjadi 98.22%, sedangkan penerapan *Bayesian Optimization* mampu meningkatkan akurasi hingga 98.34%. Dalam pendekatan prediksi, model *K-Nearest Neighbors Regressor* memperoleh nilai *MAE* sebesar 0.25001. Di sisi lain, penerapan *Grid Search* telah menurunkan nilai *MAE* menjadi 0.247545, sedangkan *Bayesian Optimization* mampu menghasilkan *MAE* yang lebih rendah dengan nilai 0.245308. Hasil penelitian ini menyimpulkan bahwa kedua metode *hyperparameter optimization* memberikan peningkatan kinerja pada kedua model konvensional. Namun, *Bayesian Optimization* secara konsisten memberikan hasil yang lebih baik daripada *Grid Search*, baik dalam meningkatkan *accuracy* untuk *Random Forest* maupun dalam mengurangi *MAE* untuk *K-Nearest Neighbors Regressor*. Dengan demikian, metode *Bayesian Optimization* telah menjadi pilihan *hyperparameter optimization* yang lebih disarankan untuk model *Random Forest* dan *K-Nearest Neighbors Regressor* dalam penelitian *Collection Intensity Scoring (CIS)* dan *channel recommendation*.

Hasil model dari *Random Forest* dan *K-Nearest Neighbors Regressor* dilakukan melalui proses prediksi dengan metode *Bayesian Optimization* sebagai model dengan kinerja terbaik. Proses prediksi ini dilakukan dengan *dummy data* yang menghasilkan visualisasi dalam bentuk tabel, *pie chart*, dan *bar plot*. Hasil model dari *Random Forest* menunjukkan prediksi kategori intensitas penagihan yang beragam pada setiap nasabah berdasarkan jumlah interaksi, durasi rata-rata interaksi menuju pembayaran, dan jumlah pinjaman yang telah dibayarkan. Selain itu, hasil model dari *K-Nearest Neighbors* menunjukkan rekomendasi *communication channel* yang tepat dengan jumlah interaksi yang diprediksi berhasil pada *channel* tersebut. Hasil kedua model ini dapat mendukung perusahaan dalam melakukan *review* dan pengambilan keputusan terkait pemberian *treatment* yang tepat pada setiap nasabah. Strategi penagihan dan komunikasi dapat disesuaikan dengan karakteristik intensitas penagihan dan perilaku interaksi nasabah sehingga mampu mendorong nasabah dalam melakukan pembayaran kembali pinjaman dengan baik.

## 5.2 Saran

Berdasarkan hasil yang diperoleh dari penelitian ini, beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya yang berkaitan dengan *Collection Intensity Scoring (CIS)* dan *channel recommendation* adalah sebagai berikut:

- 1) Meningkatkan jumlah data yang digunakan yang membantu model dalam mempelajari pola yang lebih umum untuk meningkatkan kemampuan model dalam menggeneralisasi ke data baru.
- 2) Melakukan proses *data pre-processing* yang lebih *robust*, seperti *feature engineering*, *feature scaling*, dan *feature selection* untuk mengurangi *noise* maupun kompleksitas pada *dataset*.
- 3) Menggunakan algoritma klasifikasi lainnya seperti *Decision Tree* dengan teknik *pruning* atau algoritma *boosting*, seperti *Extreme Gradient Boosting (XGBoost)* atau *Light Gradient Boosting Machine (LightGBM)* dengan kemampuan regularisasi untuk mengurangi risiko *overfitting*.
- 4) Menyesuaikan ulang model dengan mengeksplorasi berbagai rentang nilai *hyperparameter* dari metode *Grid Search* dan *Bayesian Optimization* untuk

menemukan kombinasi *hyperparameter* yang ideal demi mencapai kinerja model yang optimal.

- 5) Menerapkan *hyperparameter optimization* dengan metode optimasi lainnya, seperti *Random Search* yang menghasilkan kombinasi *hyperparameter* secara acak dengan waktu komputasi yang lebih cepat.
- 6) Melakukan *deployment* terhadap hasil klasifikasi *Collection Intensity Scoring* dan prediksi *channel recommendation* yang telah dibuat melalui *dashboard* atau *website*. Penerapan ini bertujuan agar model yang dihasilkan dapat memberikan *insight* secara *real time* guna pengambilan keputusan.

A large, light blue watermark logo of Universitas Multimedia Nusantara (UMMN) is centered on the page. It features a stylized 'U' and 'M' inside a circle, with a grid of squares below them.

# UMMN

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A