

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan zaman pada saat ini telah mendorong berbagai perubahan pada teknologi dari waktu ke waktu yang berdampak terhadap kehidupan manusia. Adapun perubahan yang berdampak signifikan berupa *Industry 4.0* dengan berbagai jenis teknologi, dimana salah satunya adalah kecerdasan buatan atau *Artificial Intelligence (AI)* yang mampu diterapkan pada berbagai jenis industri. *Artificial Intelligence* memiliki potensi kognitif yang mampu memberikan inovasi yang lebih mendalam dalam bisnis yang melibatkan pengembangan produk dan layanan baru [1]. Kehadiran *Artificial Intelligence* mampu membantu perusahaan dalam menarik perhatian pelanggan baru, namun hal tersebut tidak menutupi perusahaan dalam menghadapi sebuah tantangan. Oleh karena itu, perusahaan telah mengadopsi inovasi teknologi *Artificial Intelligence* demi beradaptasi ataupun mendisrupsi lingkungan mereka sekaligus melakukan pengembangan dan pengoptimalan keunggulan strategis maupun kompetitif bisnis [2].

Salah satu industri yang mengalami perubahan signifikan terdapat pada industri yang bergerak pada bidang keuangan. Pada awalnya, segala metode, cara, dan produk dari perbankan masih dilaksanakan dan disajikan secara tradisional. Dengan kata lain, perbankan masih menunduk pada penggunaan media kertas dalam mengurus keuangannya, serta cenderung melakukan interaksi dengan para nasabah melalui pertemuan tatap muka (*offline*). Perbankan pada umumnya seringkali membantu nasabah secara langsung dalam mengurus keuangannya sebelum kemunculan aplikasi *online*. Namun, transisi ini telah membawa perbankan menuju kelahiran perusahaan baru, yaitu perusahaan *financial technology (fintech)*. *Fintech* telah menjadi suatu tren di kalangan industri keuangan yang akan selalu berkembang dan mampu menggantikan metode keuangan tradisional dalam penyediaan layanan keuangannya. Perusahaan *fintech* dapat memperoleh pendapatan di industri ini oleh karena kapabilitas pengelolaan

teknologi berbasis *mobile* dan dipersonalisasi, serta tidak perlu mengeluarkan biaya untuk memelihara infrastruktur perbankan [3].

Salah satu fasilitas *fintech* yang kerap kali digunakan pada saat ini adalah *Peer-to-Peer (P2P) Lending*. Hal ini didukung oleh data dari Lembaga Otoritas Jasa Keuangan (OJK) pada periode September 2023 yang mencatat bahwa jumlah pengguna *fintech P2P Lending* di Indonesia telah mencapai 121.955.315 juta, serta jumlah perusahaan *fintech* berjenis *P2P Lending* di Indonesia di bawah naungan OJK telah mencapai 101 perusahaan [4]. *P2P Lending* memungkinkan adanya pertemuan yang melibatkan perjanjian akan pinjaman *online* dari pihak perusahaan terhadap pihak nasabah melalui perangkat elektronik. Dengan hal ini, nasabah hanya dapat memanfaatkan teknologi berupa komputer ataupun *smartphone* untuk mampu memperoleh pinjaman sesuai dengan kebutuhannya [5]. Oleh sebab itu, keberadaan *P2P Lending* telah memberikan kemudahan bagi kedua belah pihak dalam melakukan transaksi sehingga hal ini mengakibatkan sebagian besar masyarakat cenderung menggunakan fasilitas tersebut.

Pinjaman *online* merupakan salah satu bantuan yang dapat diberikan kepada nasabah untuk memenuhi kebutuhan hidup sehari-hari dengan metode pengajuan yang dilakukan melalui suatu *platform P2P Lending*. Dengan hal ini, para nasabah dapat memperoleh kondisi keuangan yang lebih baik, sedangkan perusahaan mampu menghasilkan pengembalian bunga yang lebih tinggi [6]. Walaupun keuntungan terdapat pada kedua belah pihak, perusahaan *P2P Lending* seringkali menghadapi berbagai permasalahan nasabah terkait pengembalian pinjaman *online*. Permasalahan nasabah yang dimaksud berupa kondisi keuangan yang buruk sehingga menimbulkan keraguan dalam hal solvabilitas dan berujung pada keterlambatan pembayaran pinjaman [7]. Nasabah sudah sepatutnya melaksanakan kewajibannya dalam memanfaatkan pinjaman dengan benar sesuai dengan perjanjian yang telah ditetapkan selama periode yang ditentukan oleh perusahaan. Oleh karena itu, perusahaan berupaya memperingati nasabah terkait pinjaman yang akan jatuh tempo. Akan tetapi, hal ini menimbulkan kesulitan karena adanya perbedaan pada perilaku nasabah dalam menanggapi peminjaman tersebut sesuai dengan kondisi keuangan dan perekonomian saat ini. Pada akhirnya, permasalahan

ini memakan biaya dan waktu yang lebih banyak bagi tenaga kerja perusahaan, terutama pihak *customer service* dalam menghadapi nasabahnya.

Perusahaan *P2P Lending* merupakan salah satu perusahaan yang bergerak di bidang *fintech* yang didirikan pada tahun 2018 dengan fokus utama pada menyediakan produk keuangan guna membantu masyarakat dalam memenuhi berbagai kebutuhan finansial. Perusahaan ini menawarkan berbagai produk pinjaman yang dapat diajukan oleh nasabah, yakni pinjaman tunai dan *paylater* yang dapat diakses melalui aplikasi *mobile*. Perusahaan *P2P Lending* berupaya untuk meningkatkan produktivitas dalam proses penagihan dan komunikasi dengan nasabahnya melalui penggunaan sistem *Collection Intensity Scoring (CIS)* sebagai salah satu produknya yang serupa dengan *credit scoring*, namun dengan fokus pada penilaian intensitas penagihan nasabah. *Credit scoring* adalah proses penghitungan atau penilaian dari *credit risk analysis* yang dimiliki ketika terdapat risiko yang dihadapi oleh pemberi pinjaman ketika memperoleh kembali investasi, terutama disebabkan oleh kegagalan peminjam dalam membayar kembali pinjaman [8]. *Credit scoring* memiliki beberapa parameter untuk menghitung skor kredit seorang konsumen, seperti *payment history*, *amounts owed*, *length of credit history*, *new credit* sebesar, dan *credit mix* [9]. Sementara itu, *CIS* telah digunakan oleh perusahaan *P2P Lending* memiliki parameter berupa *communication channel interaction*, *average interaction to payment*, dan *paid loans*. Sistem ini menggunakan data pribadi atau demografi nasabah, riwayat pembayaran pinjaman, dan interaksi melalui *communication channel* seperti *email*, *robocall*, *sms*, dan *telephony*. Dengan demikian, sistem ini melakukan penilaian dan pengelompokan intensitas penagihan nasabah menjadi intensitas penagihan rendah, sedang, dan tinggi yang mendukung perusahaan dalam berupaya memperoleh pembayaran kembali pinjaman *online* dari nasabah.

Penggunaan sistem *Collection Intensity Scoring (CIS)* pada perusahaan *P2P Lending* mengalami kendala dalam proses pengelompokan intensitas penagihan, dimana proses ini dapat memakan waktu yang lama jika dilakukan secara manual, sementara kurangnya informasi terkait *communication channel* yang digunakan oleh nasabah juga menambah kompleksitas dan biaya upaya komunikasi dari

perusahaan. Oleh karena itu, dibutuhkan model *machine learning* untuk mengklasifikasi intensitas penagihan nasabah dalam intensitas rendah, sedang, atau tinggi. Selain itu, model ini tidak hanya melakukan klasifikasi intensitas penagihan, tetapi juga memberikan *channel recommendation* serupa dengan *recommendation system* dengan prediksi jumlah interaksi nasabah sehingga membantu dalam meminimalkan biaya dan waktu yang diperlukan dalam upaya komunikasi dengan nasabah. *Recommendation system* mengambil informasi terkait kebiasaan, minat, ataupun profil pengguna dengan tujuan memberikan saran dengan informasi yang lebih sesuai atau serupa yang menarik bagi pengguna [10]. Teknik *recommendation system* dibagi menjadi 3 jenis, yakni *content-based filtering*, *collaborative filtering*, dan *hybrid*. Rekomendasi *content-based* didasarkan pada karakteristik yang mirip dengan yang disukai pengguna di masa lalu, sedangkan *collaborative filtering* didasarkan pada perilaku pengguna dari item yang direkomendasikan, dan *hybrid filtering* sebagai kombinasi dari *content-based* dan *collaborative filtering* [11]. Melalui pola perilaku interaksi pada setiap nasabah, *channel recommendation* menghasilkan daftar rekomendasi *channel* yang tepat untuk menghubungi nasabah terkait upaya pemerolehan pembayaran kembali pinjaman. Dengan *recommendation system* yang tepat sasaran, perusahaan *P2P Lending* dapat meminimalkan biaya komunikasi, serta meningkatkan produktivitas dalam proses penagihan. Dengan demikian, penelitian ini bertujuan untuk mengembangkan model *machine learning* yang dapat mengatasi permasalahan tersebut dan meningkatkan produktivitas penagihan serta komunikasi dengan nasabah.

Berdasarkan permasalahan yang telah dijabarkan, maka penelitian ini akan berfokus kepada solusi berupa *data mining*. *Data mining* dapat didefinisikan sebagai kumpulan metode untuk mengekstrak informasi yang dapat digunakan dari kumpulan data mentah yang besar [12]. Selama proses *data mining*, *Machine Learning*, statistika, dan *Artificial Intelligence (AI)* digunakan untuk memperoleh informasi tentang kemungkinan terjadinya suatu peristiwa di masa mendatang. Berbagai aspek yang dimiliki *data mining* berupa klasifikasi data, integrasi data, transformasi data, diskritisasi data, evaluasi pola, dan lain sebagainya. Metode *data mining* berguna dalam menemukan hubungan yang tersembunyi dan tidak terduga

antardata dan untuk kebutuhan pemasaran, penjualan, pendeteksi penipuan, penemuan ilmiah, pengembangan produk, perawatan kesehatan, dan pendidikan [13]. Dalam penelitian ini, penerapan *machine learning* sangat penting untuk menghasilkan model yang dapat melakukan klasifikasi *Collection Intensity Scoring (CIS)* dan prediksi *channel recommendation* pada perusahaan *P2P Lending*. Algoritma *machine learning* telah digunakan secara luas di berbagai aplikasi dan area. Namun, untuk model *machine learning* dengan masalah yang berbeda, diperlukan penyetelan pada *hyperparameter* [14]. Salah satu permasalahan yang dihadapi adalah pertumbuhan yang terus meningkat baik dalam jumlah maupun keberagaman data, penerapan metode *machine learning* konvensional menjadi tidak cukup untuk menangani permasalahan yang kompleks ini. Oleh karena itu, dibutuhkan optimisasi yang mampu meningkatkan performa model *machine learning* konvensional, yaitu *hyperparameter optimization*. *Hyperparameter optimization* merupakan proses pemilihan *hyperparameter* yang optimal untuk sebuah model *machine learning*. Optimisasi inilah sangat penting untuk memperoleh kinerja optimal dari model *machine learning* [15]. Dengan demikian, *hyperparameter optimization* berperan penting dalam penyesuaian model *machine learning* dengan perubahan lingkungan yang terus berkembang dan menghasilkan performa yang lebih optimal dalam berbagai situasi, khususnya pada model klasifikasi dan prediksi yang dilakukan pada perusahaan *P2P Lending*.

Algoritma *machine learning* yang digunakan dalam penelitian ini dipilih berdasarkan kinerja yang telah terbukti pada penelitian sebelumnya. Algoritma *Random Forest* dipilih untuk mengklasifikasi *Collection Intensity Scoring (CIS)* yang berhubungan dengan *credit scoring* karena memiliki akurasi yang tinggi dalam penelitian sebelumnya. Terdapat penelitian sebelumnya yang menggunakan algoritma *machine learning classifier* dengan *feature selection* untuk *credit scoring* untuk *German credit data*. Penelitian ini menghasilkan nilai akurasi tertinggi pada model *Random Forest* yang disertai *feature selection Chi-square* (93.12%), *Gain-Ratio* (91.20%), dan *Info-Gain* (90.90%) [16]. Selain itu, algoritma *Random Forest* pada penelitian sebelumnya menghasilkan nilai akurasi 98.4% untuk resiko gagal bayar peminjam (*default*) pada platform *P2P* di China [17]. Adapun penelitian

lainnya yang menggunakan algoritma *ensemble* untuk *dataset* pinjaman kredit konsumen *Lending Club* pada *Q4* tahun 2018, dimana algoritma *Random Forest* memiliki nilai akurasi tertinggi yang mencapai 81.05% [18].

Selain *Random Forest*, *K-Nearest Neighbors* dipilih untuk menentukan *channel recommendation* berdasarkan perilaku interaksi nasabah dan memprediksi jumlah interaksi yang akan direspon pada *channel* tersebut yang berhubungan dengan *recommendation system*. Algoritma *K-Nearest Neighbors* dipilih karena memiliki performa yang baik dalam penelitian sebelumnya dengan nilai *MAE* yang rendah. Adapun penelitian sebelumnya yang mengusulkan pendekatan *Hybrid Action-Related K-Nearest Neighbours (HAR-KNN)* yang melibatkan algoritma *KNN* untuk memberikan *product recommendation* berdasarkan riwayat pembelian, data perilaku konsumen berupa jumlah *view*, *click*, dan produk yang dibeli pada *dataset* Amazon [19]. Penelitian ini menghasilkan performa terbaik dengan nilai *MAE* sebesar 0.7165 pada 10 *neighbours* dibandingkan dengan jumlah *neighbours* dan metode lainnya. Algoritma *KNN* dengan tipe *regression (KNN regressor)* digunakan dalam penelitian ini karena menyediakan hasil rekomendasi dengan performa prediksi yang lebih baik berdasarkan *user behavior* dibandingkan dengan algoritma lainnya. Algoritma *KNN* telah memenuhi syarat secara *online* dan *real-time* untuk menemukan data *user behavior* yang mengkoordinasikan ke kelompok *user* tertentu yang berisi hubungan antara kesamaan banyak pengguna dan pengguna target dari sejumlah data yang besar [19]. Tidak hanya itu saja, penelitian selanjutnya melibatkan teknik berbasis *Machine Intelligence (MI)* untuk memberikan rekomendasi *course* dari *e-Learning* [20]. Penelitian ini menghasilkan performa terbaik dengan nilai *MAE* sebesar 0.0130, 0.875, dan 0.0136, dan 0.0142 pada model *KNN* dengan rasio *training-testing* 60:40, 70:30, dan 80:20 secara berturut-turut.

Penelitian ini melibatkan penggunaan *hyperparameter optimization* berupa *Grid Search* dan *Bayesian Optimization*. Penerapan *hyperparameter optimization* dilakukan untuk memperoleh jenis optimasi terbaik pada model *Random Forest* dan *K-Nearest Neighbors (KNN) Regressor*. *Grid search* merupakan pendekatan *hyperparameter optimization* yang sering digunakan dalam *machine learning* untuk

menemukan kombinasi *hyperparameter* yang optimal untuk model yang diberikan. Terdapat penerapan *Grid Search* dalam algoritma *ensemble* untuk kebutuhan pinjaman kredit konsumen, termasuk model *Random Forest* yang menghasilkan nilai akurasi sebesar 81.05% [18]. Dalam model *KNN*, terdapat penelitian terdahulu yang menerapkan metode *Grid Search* pada model klasifikasi dan regresi untuk memprediksi kualitas air di India, dimana model *KNN Regressor* memperoleh nilai *MAE* sebesar 0.0009 [21]. Di sisi lain, terdapat penelitian sebelumnya yang melibatkan penggunaan *Random Forest* yang dioptimasi menggunakan *Bayesian Optimization* untuk kualitas produk dalam sistem *cyber-physical* dengan nilai akurasi tertinggi sebesar 90.33% [22]. Selain itu, terdapat penerapan *KNN* dengan metode *Bayesian Optimization* untuk memprediksi erosi curah hujan di Italia dan Switzerland dengan nilai *MAE* sebesar 20.576 dalam satuan erosi curah hujan [23]. *Bayesian Optimization* mampu mencari *hyperparameter* yang optimal untuk model *machine learning* dengan mengurangi pencarian iteratif yang tidak diperlukan [24].

Penelitian ini menggunakan *Cross Industry Standard Process-Data Mining (CRISP-DM) Framework* dan proses *Extract, Transformation, Load (ETL)*. Penerapan *CRISP-DM framework* digunakan karena memiliki tahapan yang terstruktur serta mencakup pemahaman dalam aspek bisnis [25]. Selain itu, proses *ETL* digunakan dalam penelitian ini karena proses ini mendukung pengumpulan dan perubahan data dari berbagai sumber menjadi bentuk yang sesuai dan dapat digunakan untuk proses pengambilan keputusan bisnis [26].

Berdasarkan hasil penelitian sebelumnya dalam [16], [17], [18], [19], [20], [21], [22], [23] dengan berbagai algoritma *Machine Learning* untuk menghasilkan model *credit scoring* dan *recommendation system*, belum terdapat penelitian lebih lanjut mengenai pembuatan model *Random Forest* dan *K-Nearest Neighbors Regressor* untuk klasifikasi *Collection Intensity Scoring (CIS)* dan prediksi *channel recommendation* pada perusahaan *P2P Lending* dengan penerapan *Grid Search* dan *Bayesian Optimization*. Oleh karena itu, dilakukan sebuah penelitian yang berjudul “Perbandingan *Grid Search* dan *Bayesian Optimization* pada Klasifikasi *Collection Intensity Scoring* dan Prediksi *Channel Recommendation*”.

Kontribusi pada penelitian ini adalah 1) Menerapkan model klasifikasi dan prediksi dengan algoritma *Random Forest* untuk *Collection Intensity Scoring (CIS)* dan *K-Nearest Neighbors Regressor* untuk *channel recommendation*, 2) mengoptimasi model *Random Forest* dan *K-Nearest Neighbors Regressor* dengan *hyperparameter optimization* berupa *Grid Search* dan *Bayesian Optimization*, 3) melakukan perbandingan pada model *Random Forest* dan *K-Nearest Neighbors Regressor* setelah penerapan *Grid Search* dan *Bayesian Optimization* untuk menentukan model dengan jenis optimasi terbaik untuk *CIS* dan *channel recommendation* berdasarkan *accuracy* dan *Mean Absolute Error (MAE)*, 4) Mengetahui hasil pembuatan model klasifikasi *CIS* dan prediksi *channel recommendation*.

1.2 Rumusan Masalah

Berikut terdapat tiga jenis rumusan masalah yang akan dibahas dalam penelitian ini, yaitu:

- 1) Bagaimana menerapkan model *Random Forest* dan *K-Nearest Neighbors* untuk *Collection Intensity Scoring (CIS)* dan *channel recommendation*?
- 2) Bagaimana mengoptimasi model *Random Forest* dan *K-Nearest Neighbors* dengan penerapan *hyperparameter optimization (Grid Search, Bayesian Optimization)*?
- 3) Bagaimana hasil perbandingan model *Random Forest* dan *K-Nearest Neighbors* setelah diterapkan *hyperparameter optimization (Grid Search, Bayesian Optimization)* berdasarkan nilai *accuracy* dan *Mean Absolute Error (MAE)*?
- 4) Bagaimana hasil model dari *Random Forest* dan *K-Nearest Neighbors* dalam melakukan *review* atas klasifikasi *Collection Intensity Scoring* dan prediksi *channel recommendation*?

1.3 Batasan Masalah

Dalam pembuatan model terkait klasifikasi intensitas penagihan dan prediksi pada *channel recommendation*, terdapat berbagai batasan masalah yang digunakan, antara lain:

- 1) *Dataset* yang digunakan pada penelitian ini sepenuhnya milik perusahaan *P2P Lending* sebanyak 23.880 data dengan periode Juli – Desember 2023 sehingga data ini bersifat rahasia atau *confidential* dan tidak diperkenankan untuk dipublikasikan.
- 2) Penelitian ini hanya berfokus kepada perbandingan *Grid Search* dan *Bayesian Optimization* untuk *Collection Intensity Scoring (CIS)* dan *channel recommendation*, menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors Regressor*.
- 3) Penelitian ini menggunakan bahasa pemrograman Python dengan mengacu pada *framework CRISP-DM (Cross Industry Standard Process – Data Mining)*.
- 4) Proses *data preparation* pada penelitian ini melibatkan proses *ETL (Extract, Transformation, Load)* dengan bahasa pemrograman PostgreSQL (PL/pgSQL) dalam *software DBeaver*.
- 5) Proses *data mining* pada penelitian ini menggunakan *Jupyter Notebook* sebagai *Integrated Development Environment (IDE)*.
- 6) *Feature* yang digunakan untuk proses *data modeling* sebanyak 14 *feature* terkait *customer data, interaction data, loan data*.
- 7) Proses *data mining* pada penelitian ini hanya terbatas pada pembuatan model klasifikasi *Collection Intensity Scoring (CIS)* dan prediksi *channel recommendation*.
- 8) Proses *data mining* pada penelitian ini tidak membahas mengenai tahapan *deployment* karena penelitian ini menghasilkan analisis perbandingan metode optimasi pada model untuk *Collection Intensity Scoring (CIS)* dan *channel recommendation*.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berikut terdapat tujuan penelitian yang dilakukan, antara lain:

- 1) Menerapkan model *Random Forest* dan *K-Nearest Neighbors* untuk *Collection Intensity Scoring (CIS)* dan *channel recommendation*.
- 2) Mengoptimasi model *Random Forest* dan *K-Nearest Neighbors* dengan penerapan *hyperparameter optimization (Grid Search, Bayesian Optimization)*.
- 3) Mengetahui hasil model dari *Random Forest* dan *K-Nearest Neighbors* dalam melakukan *review* atas klasifikasi *Collection Intensity Scoring* dan prediksi *channel recommendation* pada *P2P Lending*.
- 4) Membandingkan hasil algoritma *Random Forest* dan *K-Nearest Neighbors Regressor* setelah diterapkan *hyperparameter optimization (Grid Search, Bayesian Optimization)* berdasarkan nilai *accuracy* dan *Mean Absolute Error (MAE)*.

1.4.2 Manfaat Penelitian

Berikut terdapat beberapa manfaat penelitian yang dilakukan, antara lain:

- 1) Manfaat Teoritis
Menambah pengetahuan terkait penggunaan *Grid Search* dan *Bayesian Optimization* pada *Random Forest* dan *K-Nearest Neighbors* untuk klasifikasi *Collection Intensity Scoring (CIS)* dan prediksi *channel recommendation*.
- 2) Manfaat Praktis
Membantu perusahaan *P2P Lending* untuk melakukan *review* atas klasifikasi intensitas penagihan nasabah dalam *Collection Intensity Scoring (CIS)* beserta prediksi *channel recommendation* sehingga mendukung perusahaan dalam menentukan keputusan terkait pendekatan atau *treatment* lebih lanjut terhadap nasabah dengan *communication channel* yang direkomendasikan.

1.5 Sistematika Penulisan

BAB I PENDAHULUAN

Dalam pendahuluan terdapat pembahasan mengenai latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas penelitian terdahulu sebagai sumber referensi dalam menunjang proses penelitian dan teori-teori umum dan khusus, seperti teori tentang *financial technology*, *P2P Lending*, *credit scoring*, *recommendation system*, *Extract, Transformation, Load (ETL)*, metode optimasi, *evaluation metrics*, *statistical tests*, *framework* dan algoritma, dan *tools* yang digunakan.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang gambaran umum objek penelitian dan metode penelitian yang digunakan, teknik pengumpulan data, variabel penelitian, dan teknik analisis data dalam penelitian.

BAB IV ANALISIS DAN HASIL PENELITIAN

Bab ini membahas tentang hasil penelitian serta pembahasan terhadap perbandingan *Grid Search* dan *Bayesian Optimization* pada model *Random Forest* dan *K-Nearest Neighbors* berdasarkan langkah-langkah penelitian yang telah ditetapkan guna menjawab rumusan masalah.

BAB V KESIMPULAN

Bab ini membahas mengenai kesimpulan dan saran dari hasil penelitian yang dilakukan.

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