

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kredit merupakan suatu kegiatan pembiayaan dari sebuah lembaga berupa pinjaman atau biasa disebut sebagai kreditur yang menyediakan dana atau sumber daya lainnya kepada nasabah sesuai dengan ketentuan yang telah ditetapkan dan disepakati oleh pihak yang bersangkutan. Istilah kata ‘kredit’ berasal dari bahasa yunani yaitu ‘*credere*’ yang berarti kepercayaan[1]. Kreditur dan nasabah biasanya membuat perjanjian yang berisikan berbagai ketentuan terkait transaksi pinjaman berupa jumlah nominal dan jangka waktu pengembalian dana. Perjanjian ini harus disepakati oleh pihak yang bersangkutan untuk menjalankan kegiatan kredit ini dengan tertib[2].

Credit Score adalah sebuah penilaian yang digunakan oleh sebuah lembaga keuangan dan kreditur untuk memperkirakan risiko dari pemberian pinjaman kepada nasabah[3]. Semakin tinggi skor kredit yang dimiliki nasabah maka, semakin baik kelayakan pinjaman yang didapatkan nasabah. Sebaliknya, apabila skor kredit yang dimiliki nasabah rendah maka, hal tersebut dapat menyulitkan nasabah untuk mendapatkan pinjaman[4]. Penentuan *Credit Score* ini diambil dari berbagai macam transaksi yang dilakukan oleh nasabah termasuk ketertiban nasabah dalam melakukan transaksi pinjaman. Apabila nasabah melakukan transaksi pinjaman tepat waktu atau bahkan sebelum mencapai jangka waktu yang telah ditentukan maka *Credit Score* yang didapatkan nasabah tinggi. Sebaliknya, jika nasabah telat melakukan transaksi pinjaman atau melewati jangka waktu yang telah ditentukan maka *Credit Score* yang didapatkan nasabah akan rendah [5].

Penentuan *Credit Score* dapat dilakukan dengan memanfaatkan teknik *Machine Learning*. Namun, salah satu tantangan utama dalam proses ini adalah banyaknya fitur pada data yang digunakan untuk penentuan *Credit Score*. Beberapa dari fitur tersebut mungkin tidak memiliki relevansi terhadap risiko kredit, sementara fitur yang lain mungkin mengalami rendundan dikarenakan adanya interkorelasi pada data yang tinggi sehingga mengakibatkan penurunan akurasi model prediksi untuk *Credit Scoring* serta banyaknya waktu yang dibutuhkan dalam komputasi karena

besarnya data yang harus diproses oleh sistem [6]. Oleh karena itu, perlu dilakukan proses seleksi fitur pada data guna mengatasi permasalahan tersebut.

Untuk memilih fitur – fitur yang independen, memiliki prediktivitas tinggi, dan mengurangi redundansi, terdapat beberapa teknik yang dapat diimplementasikan. Salah satunya adalah dengan menggunakan metode analisis korelasi statistik[7]. Proses seleksi fitur juga dapat melibatkan penerapan *machine learning* yang digunakan untuk memilih fitur-fitur yang paling relevan dalam membuat prediksi. Dengan mengintegrasikan metode korelasi statistic dan algoritma *machine learning* sehingga dapat menghasilkan model *Credit Scoring* yang lebih efisien, akurat, dan memiliki interpretasi tinggi.

Adapun penelitian sebelumnya yang terkait dilakukan oleh Jin Hee Bae, Minwoo Kim, J.S. Lim, dan Zong Woo Geem dengan judul “*Feature Selection for Colon Cancer Detection Using K-Means Clustering and Modified Harmony Search Algorithm*”. Tujuan dari penelitian ini adalah untuk mengembangkan metode seleksi fitur yang dapat membedakan antara pasien penderita kanker kolorektal dari individu normal dengan menggunakan metode K-Means dan algoritma Harmony Search yang telah dimodifikasi. Hasil yang didapat dengan menggunakan metode tersebut mencapai akurasi klasifikasi hingga 94,36%. Selanjutnya, penelitian yang dilakukan oleh Juan Laborda dan Soyeong dengan judul “*Feauture Selection in a Credit Scoring Model*”. Tujuan dari penelitian ini adalah untuk meningkatkan model credit score dengan menfaatkan machine learning dan data non-tradisional. Metode yang digunakan dalam penelitian ini meliputi *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Random Forest*. Hasilnya menunjukkan bahwa pemilihan fitur yang relevan secara signifikan dapat mempengaruhi kinerja model credit score.

Berdasarkan penelitian sebelumnya, maka dalam penelitian ini mengusulkan seleksi fitur pada penentuan *credit score* yang mengimplementasikan metode *K-Means* dan *Harmony Search*. Tujuannya adalah untuk mencari kombinasi fitur yang paling relevan dan informatif dalam menentukan *credit score*. Dalam penelitian ini analisis korelasi antar fitur tidak dilakukan secara langsung menggunakan koefisien korelasi statistik, melainkan melalui pendekatan *distance-based* menggunakan metode *K-Means*. Metode ini akan melakukan proses pengelompokan pada data

dan mengidentifikasi pola-pola yang ada dalam data. Hasil dari pengolompokkan ini dapat dilakukan analisis korelasi secara tidak langsung menggunakan hasil perhitungan jarak *Euclidean Distance* dan melakukan proses pemilihan fitur yang paling representatif sebelum melalui proses selanjutnya. Sedangkan *Harmony Search* digunakan untuk mencari kombinasi fitur yang relevan untuk meningkatkan prediksi *credit scoring*.

1.2. Rumusan Masalah

Berdasarkan dengan latar belakang yang telah dijelaskan diatas, maka peneliti dapat merumuskan permasalahan dari penelitian ini sebagai berikut :

1. Bagaimana penerapan K-Means dalam melakukan pengelompokkan data untuk mendukung proses seleksi fitur pada penentuan *credit score* ?
2. Bagaimana penerapan Harmony Search dalam proses seleksi fitur pada penentuan *credit score* ?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan diatas, tujuan dari penelitian ini adalah untuk menganalisis penerapan metode *K-Means* dalam mendukung proses seleksi fitur pada penentuan *Credit Score* dan menganalisis penerapan metode *Harmony Search* dalam menghasilkan kombinasi fitur terbaik pada seleksi fitur. Serta mengevaluasi kualitas data yang dihasilkan berdasarkan fitur-fitur yang terpilih dengan mengukur tingkat kesalahan atau *eror* dan melihat efisiensi waktu komputasi.

1.4. Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini, sebagai berikut:

1. Menemukan kombinasi fitur yang baik sehingga dapat meningkatkan kualitas data dalam penentuan credit score.
2. Mengevaluasi efektivitas algoritma K-Means dan Harmony Search dalam melakukan seleksi fitur pada penentuan credit score.

1.5. Batasan Masalah

Dalam penelitian ini, penulis membatasi penelitian supaya mendapatkan hasil yang diinginkan.

1. Data dalam penelitian ini diambil dari kaggle dengan nama data “Credit Score”.

2. Topik pembahasan dari penelitian ini dibatasi sampai dengan seleksi fitur dan prediksi eror yang didapat dari fitur yang terpilih.
3. Menggunakan bahasa pemrograman python pada penelitian ini.