BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Penerapan teknik *machine learning* untuk deteksi anomali telah berkembang di berbagai bidang, seperti keuangan, keamanan, dan manufaktur. Dalam beberapa tahun terakhir, pertumbuhan data proses bisnis serta perkembangan metode analisis data mendorong meningkatnya penelitian mengenai deteksi anomali berbasis *event log* [1]. Hal ini relevan pada sektor perbankan karena proses bisnis perbankan memiliki volume data yang besar, regulasi yang ketat, serta alur proses yang kompleks. Salah satu proses penting dalam perbankan adalah proses pengajuan kredit, yang melibatkan tahapan evaluasi, validasi, komunikasi, dan pengambilan keputusan. Penyimpangan pada proses tersebut dapat mengindikasikan potensi keterlambatan, pengulangan pekerjaan, alur yang tidak efisien, atau kebutuhan eskalasi dalam proses bisnis [2]. Oleh karena itu, deteksi anomali pada proses pengajuan kredit dapat digunakan sebagai sistem deteksi awal (*early detection system*) untuk membantu mengidentifikasi kasus yang memerlukan analisis lebih lanjut [3].

Data proses bisnis umumnya direpresentasikan dalam bentuk *event log*, yang memuat informasi mengenai identitas kasus, aktivitas, waktu kejadian, dan sumber daya yang terlibat. Karakteristik ini memungkinkan analisis terhadap perilaku proses berdasarkan urutan aktivitas, durasi, keterlibatan sumber daya, serta variasi alur proses. Namun, dalam praktiknya data *event log* sering kali tidak memiliki label anomali aktual, sehingga pendekatan *supervised learning* sulit diterapkan secara langsung. Kondisi ini menjadi tantangan utama dalam penelitian deteksi anomali, karena model tetap perlu mengidentifikasi kasus yang berpotensi menyimpang meskipun tidak tersedia label dari pakar atau sistem operasional.

Sejumlah penelitian menunjukkan bahwa berbagai pendekatan *machine learning* telah digunakan dalam deteksi anomali. Pendekatan *supervised learning*, seperti *Support Vector Machine* (SVM), memiliki kemampuan dalam membangun batas keputusan untuk membedakan kelas normal dan anomali, tetapi bergantung pada ketersediaan data berlabel [4][5]. Di sisi lain, pendekatan *unsupervised learning*, seperti *clustering*, dapat digunakan pada data tanpa label dan mampu mengidentifikasi pola kepadatan serta *outlier* [6][7]. DBSCAN merupakan salah

satu metode *clustering* berbasis kepadatan yang dapat mengidentifikasi data sebagai *noise* apabila tidak termasuk ke dalam kelompok data yang padat. Temuan pada [4] menunjukkan bahwa DBSCAN efektif untuk mendeteksi anomali berbasis kepadatan, sedangkan SVM memiliki keunggulan dalam proses klasifikasi. Selain itu, metode lain seperti Isolation Forest dan pendekatan berbasis *deep learning* juga telah digunakan untuk mendeteksi anomali dengan karakteristik yang berbeda [4][8]. Hal ini menunjukkan bahwa masing-masing metode memiliki kelebihan dan keterbatasan, sehingga pemilihan metode perlu disesuaikan dengan karakteristik data dan tujuan penelitian.

Keterbatasan ketersediaan label pada data proses bisnis mendorong pengembangan berbagai pendekatan yang memanfaatkan *pseudo-label* sebagai alternatif dalam proses pembelajaran model. Salah satu pendekatan yang banyak diusulkan adalah kombinasi antara metode *clustering* dan *supervised learning*, di mana hasil pengelompokan digunakan untuk menghasilkan *pseudo-label* yang selanjutnya dimanfaatkan dalam proses klasifikasi [9, 10, 11]. Konsep tersebut terinspirasi dari mekanisme *pseudo-labeling* yang banyak digunakan pada pendekatan *semi-supervised learning* dan *weakly supervised learning*, yaitu memanfaatkan label sementara untuk membantu proses pembelajaran ketika label aktual tidak tersedia atau sangat terbatas [12, 13]. Pada penelitian ini, konsep tersebut diadaptasi melalui penggunaan algoritma DBSCAN untuk menghasilkan *pseudo-label* berdasarkan hasil *clustering*, kemudian SVM digunakan untuk mempelajari pola pemisahan antara data normal dan kandidat anomali berdasarkan *pseudo-label* tersebut.

Berdasarkan konsep tersebut, penelitian ini menerapkan kombinasi algoritma DBSCAN dan SVM untuk mendeteksi kandidat anomali pada data *event log* proses pengajuan kredit perbankan. DBSCAN digunakan untuk mengidentifikasi *noise* sebagai kandidat anomali sekaligus menghasilkan *pseudo-label*, sedangkan SVM digunakan untuk mempelajari pola klasifikasi berdasarkan *pseudo-label* tersebut. Pendekatan ini tidak menggunakan label anomali aktual dalam proses pelatihan, sehingga berfungsi sebagai mekanisme deteksi awal (*early anomaly detection*) yang membantu mengidentifikasi kasus-kasus yang memerlukan analisis lebih lanjut. Oleh karena itu, evaluasi model difokuskan pada kemampuan mendeteksi kandidat anomali, khususnya melalui analisis *confusion matrix*, *precision*, *recall*, dan *F1-score*, dengan penekanan pada nilai *recall* untuk meminimalkan risiko anomali yang tidak terdeteksi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini dirumuskan dalam beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana mengimplementasikan pendekatan berbasis DBSCAN dan SVM untuk mendeteksi kandidat anomali pada proses pengajuan kredit perbankan dengan kondisi data tanpa label aktual?
2. Bagaimana karakteristik kasus yang terdeteksi sebagai anomali berdasarkan fitur-fitur *event log* pada proses pengajuan kredit perbankan?
3. Bagaimana kinerja pendekatan DBSCAN-SVM berbasis *pseudo-labeling* dalam mendeteksi anomali pada data tanpa label berdasarkan *confusion matrix*?

1.3 Batasan Permasalahan

Untuk menjaga fokus penelitian diperlukan batasan terhadap ruang lingkup permasalahan. Batasan ini bertujuan untuk memperjelas cakupan penelitian sehingga proses analisis dan interpretasi hasil dapat dilakukan secara lebih terarah dan sistematis. Adapun batasan permasalahan dalam penelitian ini adalah sebagai berikut.

1. Penelitian ini dilakukan menggunakan data sekunder berupa *event log* proses pengajuan kredit perbankan yang berasal dari dataset *Business Process Intelligence Challenge 2017*. Dataset tersebut merepresentasikan proses bisnis pada institusi keuangan di Belanda dan tidak melibatkan pengambilan data langsung dari sistem operasional nyata.
2. Ruang lingkup penelitian dibatasi pada deteksi kandidat anomali pada proses pengajuan kredit perbankan (*loan application process*) berdasarkan data *event log*.
3. Data yang digunakan tidak memiliki label anomali aktual, sehingga label yang digunakan dalam proses pemodelan merupakan *pseudo-label* yang dibentuk dari hasil clustering DBSCAN.

4. Pendekatan yang digunakan difokuskan pada kombinasi DBSCAN dan SVM. Metode lain hanya dibahas secara konseptual dan tidak dibandingkan secara eksperimental.

5. Hasil deteksi anomali pada penelitian ini digunakan sebagai deteksi awal (*early detection*) dan dasar prioritas analisis lebih lanjut, bukan sebagai keputusan akhir terhadap status suatu kasus.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan sebelumnya, tujuan dari penelitian ini adalah sebagai berikut.

1. Mengimplementasikan pendekatan berbasis DBSCAN dan SVM untuk mendeteksi kandidat anomali pada proses bisnis pengajuan kredit perbankan dengan kondisi data tanpa label aktual, melalui pembentukan *pseudo-label* dari hasil *clustering* DBSCAN.

2. Mengevaluasi kinerja pendekatan yang diusulkan berdasarkan *confusion matrix*, *precision* dan *recall*, dengan penekanan pada nilai *recall* karena penelitian ini berfokus pada deteksi awal anomali dan upaya meminimalkan kasus anomali yang tidak terdeteksi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat antara lain.

1. Memberikan pendekatan deteksi awal (*early detection*) terhadap kandidat anomali pada proses pengajuan kredit perbankan berbasis data *event log* yang tidak memiliki label aktual.

2. Membantu mengidentifikasi kasus yang berpotensi memerlukan analisis lebih lanjut, terutama kasus yang berkaitan dengan durasi proses yang panjang, pengulangan aktivitas, keterlibatan banyak sumber daya, atau alur proses yang tidak efisien.

3. Memberikan kontribusi pemahaman mengenai penerapan berbasis DBSCAN dan SVM melalui pemanfaatan *pseudo-label* untuk deteksi anomali pada data proses bisnis tanpa label.

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan laporan penelitian yang menjelaskan secara ringkas isi dari setiap bab yang disusun dalam penelitian ini:

1. Bab 1 PENDAHULUAN

   Berisi latar belakang penelitian, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan laporan.

2. Bab 2 LANDASAN TEORI

   landasan teori dan tinjauan pustaka yang relevan dengan penelitian. Bab ini juga menyajikan hasil studi literatur terkait penelitian terdahulu.

3. Bab 3 METODOLOGI PENELITIAN

   Menjelaskan metodologi penelitian yang digunakan, meliputi deskripsi data, tahapan prapemrosesan data, perancangan model, skema pembagian dataset, serta mekanisme validasi.

4. Bab 4 HASIL DAN DISKUSI

   Menyajikan hasil implementasi, termasuk hasil evaluasi performa model berdasarkan metrik, serta pembahasan mengenai efektivitas pendekatan yang diusulkan.

5. Bab 5 KESIMPULAN DAN SARAN

   Berisi simpulan dari keseluruhan hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.