BAB I
PENDAHULUAN

1.1. Latar Belakang

Persediaan atau stok memegang peran penting dalam keberlangsungan proses bisnis karena sangat erat kaitannya dengan arus kas sebuah perusahaan (Zhong & Zhang, 2015). Perusahaan dengan model bisnis jual beli harus memastikan bahwa barang yang akan dijual telah dimiliki dalam bentuk persediaan. Kegiatan pengendalian persedian sering kali berdampingan dengan permintaan berselang yang memiliki variabilitas ukuran dan waktu antar permintaan sehingga dapat menjadi masalah dalam kegiatan jual beli barang. Penjualan barang pada umumnya sangat dipengaruhi oleh kelengkapan dan kuantitas barang yang dimiliki oleh sebuah perusahaan, menjadikan keputusan pembelian dalam upaya pengendalian persediaan barang adalah hal yang krusial.

Permintaan barang yang berselang merupakan permasalahan bagi proses pengendalian persediaan, hal ini menjadi tantangan untuk dapat diselesaikan. Seperti penelitian yang dilakukan oleh Pennings dkk. (2016) melakukan peramalan untuk permintaan yang berselang (terputus-putus) bersumber dari data permintaan yang lalu bertujuan untuk mengantisipasi lonjakan permintaan yang masuk di masa mendatang. Penelitian mengenai permintaan dan manajemen persediaan juga telah banyak dikembangkan. Pada penelitian yang dilakukan oleh Zhong & Zhang (2015) dikatakan bahwa para praktisi telah menerapkan berbagai metode untuk menentukan level dari stok aman, diantaranya menggunakan metode ABC, perangkingan, marginal analisis, matematika statistik dan masih banyak lagi.

Dari penerapan berbagai metode yang ada, klasifikasi dengan multikerteria menjadi yang paling banyak digunakan (Bacchetti & Saccani, 2012). Penelitian sebelumnya yang menggunakan data publik yang sama, yaitu mengenai prediksi untuk pembelian kembali sebuah barang diantaranya dilakukan oleh Li (2017). Li (2017) memprediksi pembelian kembali pada studi kasus nyata *Danish Craft Beer Breweries* untuk membandingkan kinerja *Supply Chain Management* tradisonal dengan metode yang disarankan, yaitu k-NN, SVM, *Logistic Regression*, dan *Decision Tree*. Penelitian Li (2017) menghasilkan nilai akurasi beragam untuk

setiap model dengan perlakuan awal data dan metode yang berbeda. Pada proses perlakuan awal data dilakukan penyeimbangan data dengan metode *Random Under-Sampling* (RUS), penghilangan *outliers* dengan teknik *Standar Deviation*, penskalaan data dengan metode *Min-Max Normalization*, dan reduksi atribut data dengan metode PCA. Kesimpulan yang didapatkan pada penelitian Li (2017) adalah SVM memiliki nilai peforma yang paling stabil dan akurasi prediksi terbaik. Sehingga dari keseluruhan pengujian didapatkan SVM *Cubic Kernel* dengan perlakuan awal data tanpa *outliers* pada model atribut hasil reduksi PCA adalah hasil yang terbaik dengan nilai akurasi 86%.

Selanjutnya penelitian yang dilakukan oleh Pritalia (2018) yang mengusulkan metode klasifikasi C45 untuk menganalisa waktu pembelian barang kembali pada *e-commerce* dengan hasil keputusan klasifikasi adalah melakukan pembelian kembali atau tidak. Data penelitian berjumlah 5.000 baris data yang dipilih secara acak dengan 23 atribut menggunakan tools software Weka. Hasil dari penelitian ini menghasilkan tingkat akurasi sebesar 98.9%.

Pada penelitian yang dilakukan oleh Hajek & Abedin (2020) melakukan pemaksimalan keuntungan untuk inventori barang menggunakan metode k-NN, *Logistic Regression*, NN, SVM, RF dan XGBoost. Hajek & Abedin (2020) menggunakan dasar *metode k-means* untuk berbagai metode perlakuan awal data dan mengusulkan penanganan terhadap data yang tidak seimbang menggunakan metode CBUS (*Classification-based Under-sampling*). Pada penelitian yang dilakukan oleh Hajek & Abedin (2020) menghasilkan akurasi tertinggi dengan presentasi 92% menggunakan metode RF.

Dari penelitian yang telah dilakukan menggunakan metode klasifikasi dan data publik yang sama didapatkan nilai presentase akurasi yang berbeda-beda. Hal ini dapat disebabkan oleh perlakuan awal data (*pre-processing*) yang berbeda-beda untuk setiap penelitian sebelumnya. Pengolahan awal terhadap data sangat diperlukan, selain memastikan data layak dan baik digunakan dalam pengujian, pengolahan awal data juga sangat berperan dalam menentukan tingkat akurasi dan presisi hasil pengujian (Larose, 2014:17). Pada penelitian ini akan dilakukan serangkaian pengolahan awal data untuk mengoptimalkan hasil klasifikasi model yang diusulkan. Proses penghapusan *outliers* dengan menggunakan metode

Interquartile Range Method diusulkan untuk memberikan peningkatan nilai akurasi dan presisi prediksi klasifikasi (Ilyas & Chu, 2019:12). Reduksi dimensi data dengan metode *Principal Component Analysis* juga dilakukan untuk efisiensi komputasi dalam proses pembelajaran dan pengujian model. Analisa dampak yang dihasilkan dari reduksi atribut data terhadap penurunan nilai akurasi yang dihasilkan menjadi hal utama untuk ditinjau (Larose, 2014:17). Data yang digunakan pada penelitian ini merupakan data publik pada situs Kaggle dengan judul 'Transaction Data with Back-order'. Data tersebut sangatlah tidak seimbang, sehingga dilakukan *undersampling* data dengan menggunakan metode *Random Undersampling*. Teknik *undersampling* dapat menghindarkan model dari generalisasi hasil prediksi klasifikasi (Brownlee, 2020). Hal ini penting dilakukan untuk mendapatkan *F1 Score* dan nilai akurasi yang baik.

Pada penelitian ini dilakukan klasifikasi untuk memprediksi *backorder* dengan dua keputusan, yaitu 'iya' untuk melakukan *backorder* dan 'tidak' untuk tidak melakukan *backorder,* berbasis *Support Vector Machine* (SVM) dan *Random Forest* (RF). Pemilihan metode SVM karena dianggap sesuai dengan data kasus yang memiliki dua label keputusan, yaitu 'iya' untuk melakukan *backorder* dan 'tidak' untuk tidak melakukan *backorder,* sesuai dengan metode SVM yang terbukti baik dalam *binary classification*, dengan berbagai keunggulan lain seperti *maximum margin classifier* dan *powerfull non-linear approximator* (Kowalczyk, 2017). Pemilihan metode *Random Forest* didasari bahwa RF merupakan metode *ensemble*, yaitu metode yang menggunakan kombinasi dari beberapa model, dengan menerapkan metode *bootstrap aggregating* dan *random feature selection*. *Bootstrap aggregating* atau *bagging* digunakan untuk memperbaiki hasil dari algoritma klasifikasi juga mengurangi *varians* dan membantu menghindari *overfitting* (Brownlee, 2016).

Pada penelitian ini dilakukan proses perlakuan awal data dengan runtutan dan metode yang berbeda dari penelitian sebelumnya. Pengolahan awal data yang lebih bervariasi diharapkan dapat menangani permasalahan dan meningkatkan kualitas hasil pengujian data. Pada penelitian ini juga dilakukan pembangunan dua jenis model yang berbeda, yaitu model dengan atribut lengkap dan model dengan atribut hasil reduksi metode PCA dengan dan tanpa penghapusan *outliers*. Perlakuan awal

data yang baik diharapkan menghasilkan model yang dapat digunakan untuk mengurangi biaya persediaan, meningkatkan daya saing pasar, dan memenuhi kebutuhan pelanggan dengan lebih baik pada perusahaan sejenis di masa yang akan datang.

1.2. Rumusan Masalah

1. Bagaimana proses pengolahan awal data untuk menangani permasalahan yang dimiliki oleh data publik dari situs Kaggle dengan judul 'Transaction Data with Back-order'?
2. Bagaimana membangun model atribut lengkap dan model atribut hasil reduksi metode PCA dengan dan tanpa penghapusan *outliers* untuk data publik dari situs Kaggle dengan judul 'Transaction Data with Back-order'?
3. Bagaimana mengimplementasikan metode klasifikasi *Support Vector Machine* (SVM) dan *Random Forest* (RF) untuk melakukan prediksi *backorder* pada data 'Transaction Data with Back-order'?

1.3. Tujuan

1. Melakukan proses pengolahan awal data untuk menangani permasalahan yang dimiliki oleh data publik dari situs Kaggle dengan judul 'Transaction Data with Back-order'.
2. Membangun dua model data yang berbeda, yaitu model dengan atribut lengkap dan model dengan atribut hasil reduksi metode PCA dengan dan tanpa penghapusan *outliers* untuk data publik dari situs Kaggle dengan judul 'Transaction Data with Back-order'.
3. Mengimplementasikan metode klasifikasi *Support Vector Machine* (SVM) dan *Random Forest* (RF) untuk melakukan prediksi *backorder* pada dua model data yang diusulkan sebagai upaya pengendalian persediaan barang.
4. Menemukan dan membandingkan peforma *overall accuracy* untuk pengujian dua model yang diusulkan, yaitu model atribut lengkap dan model atribut hasil reduksi metode PCA dengan dan tanpa penghapusan *outliers* menggunakan metode klasifikasi *Support Vector Machine* (SVM) dan *Random Forest* (RF).

1.4. Manfaat

1. Penerapan ilmu pengetahuan yang diperoleh dari perkuliahan.
2. Proses pengolahan awal data yang optimal untuk pengembangan sistem prediksi *backorder* perusahaan sejenis di waktu yang akan datang.
3. Bagi akademisi sebagai referensi untuk melakukan penelitian lebih lanjut terkait proses pengolahan awal data dan prediksi *backorder* untuk pengendalian persediaan barang berbasis *Support Vector Machine* (SVM) dan *Random Forest* (RF).

1.5. Batasan Masalah

1. Data yang digunakan pada penelitian ini adalah data publik pada situs Kaggle dengan judul ‘Transaction Data with Back-order’ yang dapat diakses melalui https://www.kaggle.com/mushfique917/transaction-data-with-backorder.
2. Data yang dipilih dalam penelitian ini terdiri 1.929.935 baris data dengan 22 atribut dan 1 atribut keputusan.
3. Penelitian ini melakukan *binary classification* untuk target keputusannya, yaitu keputusan ‘iya’ untuk melakukan *backorder* dan ‘tidak’ untuk tidak melakukan *backorder*.