

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Konsep data mining banyak diterapkan dalam berbagai bidang ilmu, seperti bidang kedokteran, analisis pemasaran, pendidikan dan rekayasa manufaktur. Data mining dapat digunakan untuk melakukan proses klasifikasi, prediksi, perkiraan dan memperoleh informasi yang bermanfaat dari sekumpulan data berukuran besar. Data mining merupakan suatu penemuan informasi dengan melakukan penggalian pola informasi yang berisi pencarian trend dalam sejumlah data yang sangat besar serta membantu proses penyimpanan data dalam mengambil suatu keputusan diwaktu yang akan datang[1]. Dalam data mining fungsi merupakan salah satu bagian dari fungsi minor, salah satunya adalah prediksi. Prediksi adalah cara untuk mencari kemungkinan hasil pada masa akan datang[2].

Seiring dengan tingginya aktivitas dan bisnis, mobil sudah menjadi kebutuhan pokok. Disisi lain, harga mobil baru semakin tinggi dengan berbagai *feature* yang disematkan dalam produk baru. Untuk memenuhi kebutuhan tersebut, masyarakat mencari alternatif dengan membeli mobil bekas[3]. Banyak hal yang harus diperhatikan ketika membeli sebuah mobil bekas, hal tersebut dikarenakan mobil telah digunakan sebelumnya oleh orang lain dan nilai kualitas barangnya tentunya mengalami penyusutan, salah satu faktor yang harus dipertimbangkan juga adalah masalah harga[4]. Dalam hal ini peneliti akan melakukan pengujian terhadap dua algoritma yang cukup populer dalam hal prediksi yaitu, algoritma *Support Vector Machine* dan algoritma *Linear Regression* dalam prediksi harga mobil bekas.

Support Vector Machine merupakan metode yang termasuk kedalam *supervised learning* yang menganalisa data dan mengenali pola untuk klasifikasi[5]. *Support Vector Machine* memiliki kemampuan untuk mengatasi masalah yang bersifat linear dan nonlinear. Pendekatan kernel yang dikembangkan pada SVM juga dapat digunakan untuk mengatasi jumlah kelas yang beragam dan banyak, SVM untuk prediksi disebut dengan *SVM Regression* terdiri dari fungsi linear dan nonlinear[6]. Untuk algoritma berikutnya yaitu algoritma *Linear Regression*, algoritma ini merupakan pemodelan dan analisis data numerik yang terdiri dari satu atau lebih variabel independen dan nilai variabel dependen[7]. Hubungan variabel dependen dan independen tergantung dalam beberapa bentuk persamaan, sebagai berikut hubungan

linear, eksponensial dan yang terakhir berganda, dengan tujuan penggunaan analisis regresi adalah untuk mengestimasi nilai variabel dependen yang didasarkan pada nilai variabel independen[8].

Beberapa penelitian yang pernah dilakukan sebelumnya mencakup “Analisis dan Perbandingan Algoritma Data Mining dalam Prediksi Harga Saham GGRM”, dimana diperoleh hasil dari perbandingan algoritma *Neural Network*, *Linear Regression*, *Support Vector Machine*, *Gaussian Process*, dan *Polynomial Regression*, disimpulkan bahwa data harga saham GGRM dapat diprediksi menggunakan model algoritma *Neural Network* dengan hasil akurasi prediksi RMSE 612.474 +/- 89.402 (mikro : 618.916 +/- 0.000) paling kecil dibandingkan dengan model algoritma lainnya, sehingga dengan prediksi ini dapat membantu dalam memprediksi harga saham GGRM di pasar modal[9]. Penelitian lainnya mencakup “Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode *Naïve Bayes* dan *Support Vector Machine (SVM)*”, dimana hasil penelitian yang diperoleh adalah persentase akurasi tertinggi diperoleh *SVM Polynomial*, sedangkan *Naïve Bayes* menghasilkan persentase akurasi terendah[10]. Penelitian lainnya mencakup “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter” dimana hasil penelitian diperoleh perhitungan akurasi yang menunjukkan bahwa metode *Support Vector Machine* lebih unggul dengan nilai 89,70% tanpa *K-Fold Cross Validation* dan 88,76% dengan *K-Fold Cross Validation*. Sedangkan pada perhitungan waktu proses metode *K-Nearest Neighbor* lebih unggul dengan waktu proses 0,0160s tanpa *K-Fold Cross Validation* dan 0,1505s dengan *K-Fold Cross Validation*[11]. Penelitian lainnya mencakup “Perbandingan performa Algoritma *Neural Network*, Regresi Linear, dan *Random Forest* dalam simulasi prediksi angka kematian pasien COVID-19 di Indonesia”, dimana hasil penelitian diperoleh bahwa algoritma *random forest* adalah algoritma yang paling mendekati nilai asli dalam melakukan prediksi dengan nilai RMS 94,991 +/- 0,000 selanjutnya diikuti oleh algoritma *neural network* dengan nilai RMS 285,956 +/- 0,000 simpangan terjauh terjadi pada algoritma *linear regression* yaitu *positive* 3285 menyimpang dari nilai sebenarnya[12].

Untuk mengetahui kinerja dari kedua algoritma tersebut dalam melakukan prediksi harga mobil bekas, maka diperlukan proses analisis kinerja dari kedua algoritma tersebut. Berdasarkan uraian tersebut, maka judul penelitian yang diangkat

yaitu “**Analisis Perbandingan Algoritma SVM Dengan Linear Regression Dalam Memprediksi Harga Mobil Bekas**”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka yang menjadi permasalahan adalah bagaimana menganalisis kinerja dari algoritma SVM dengan algoritma *linear regression* dalam melakukan proses memprediksi harga mobil bekas.

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Tujuan dari penelitian untuk membandingkan kedua algoritma ini adalah untuk menentukan algoritma mana yang menghasilkan nilai akurasi yang lebih baik dalam hal memprediksi harga sebuah mobil bekas.

1.3.2 Manfaat

Adapun manfaat yang diperoleh dengan selesainya penelitian ini :

1. Mengetahui perbandingan kinerja dari algoritma SVM dan *Linear Regression*.
2. Mengetahui nilai akurasi mana yang lebih baik dari kedua algoritma yang diuji.
3. Menjadi bahan referensi untuk pembelajaran dan penelitian dalam proses memprediksi.

1.4 Keterbaruan

Berikut merupakan kumpulan penelitian – penelitian terdahulu yang peneliti jadikan sebagai bahan referensi dalam penyusunan proposal penelitian peneliti :

1. “Implementasi Data Mining Menggunakan Model SVM Untuk Prediksi Kepuasan Pengunjung Taman Tabebuya”, dimana hasil penelitian ini dapat disimpulkan bahwa algoritma SVM memiliki akurasi dan performa secara rata – rata yaitu sebesar 86,00% dan nilai AUC sebesar 0.947[13].
2. “Prediksi Waktu Kelulusan Mahasiswa Menggunakan SVM Berbasis PSO”, dimana hasil penelitian yang didapat yaitu, dengan menggunakan model *support vector machine* berbasis *particle swarm optimization* dapat meningkatkan akurasi prediksi dari sebesar 85,81% menjadi 86,43% dengan kenaikan sebesar 00.62%. Sehingga dalam memprediksi kelulusan mahasiswa

dapat akurat dan secara optimal dalam mengukur parameter yang diperlukan[14].

3. “Analisis dan Penerapan Algoritma *Support Vector Machine* (SVM) dalam Data Mining Untuk Menunjang Strategi Promosi”, hasil penelitian ini menggunakan riset data rekapitulasi penerimaan mahasiswa baru prodi informatika UPGRIS dari tahun 2014 sampai 2018 dengan total 6371 baris. Berdasarkan hasil analisis diperoleh tingkat akurasi sebesar 73,6%[15].
4. “Implementasi Jaringan Syaraf Tiruan Metode *Backpropagation* untuk Prediksi Penjualan Mobil Bekas”, hasil penelitian dalam penelitian ini menunjukkan bahwa sistem aplikasi yang dibuat dapat menghasilkan hasil prediksi yang cukup akurat untuk mendapatkan hasil yang tidak jauh berbeda dengan penjualan aktual yaitu dengan nilai $MSE=23,84$ [16].
5. “Implementasi Metode *K-Nearest Neighbor* Dalam Peramalan Penjualan Mobil Bekas Di Kota Batam”, hasil penelitian dalam penelitian ini adalah penelitian ini menggunakan dataset yang diambil dari tahun 2017 sampai 2020, setelah di uji menggunakan algoritma K-NN dengan bantuan *software rapidminer*, tingkat akurasi yang didapat adalah 63,89% dengan level presisi 100%[17].
6. “Klasifikasi Penerimaan Mobil Bekas Berdasarkan Metode *Neural Network*”, penelitian ini menggunakan metode jaringan syaraf tiruan dengan data sebanyak 1728 data yang diperoleh dari UCI *repository* dengan tingkat akurasi yang cukup tinggi yaitu sebesar 98.26%[18].
7. “Prediksi Kebutuhan Alat Kesehatan Rumah Sakit Menggunakan Metode Algoritma *Regression Linear* dan *Naïve Bayes*”, kesimpulan dari penelitian ini bahwa hasil prediksi alat kesehatan di RSUD Pali berdasarkan dataset yang diimplementasikan dengan metode algoritma *regression linear* dan *Naïve Bayes* menunjukkan hasil bahwa prediksi algoritma *regression linear* memiliki tingkat nilai akurasi yang lebih tinggi[19].
8. “Prediksi Harga Rumah Menggunakan Web *Scrapping* dan *Machine Learning* Dengan Algoritma *Linear Regression*”, kesimpulan dari penelitian ini adalah untuk mendapatkan nilai prediksi yang tinggi penelitian dilakukan secara berulang kali namun nilai prediksi terbesar dengan menggunakan 80% dataset

untuk training dan 20% dataset digunakan untuk testing menghasilkan nilai output dengan tingkat akurasi dalam memprediksi sebesar 88%[20].

Berdasarkan penelitian – penelitian terdahulu maka pada penelitian kali ini dilakukan dengan melakukan perbandingan dua algoritma yaitu algoritma svm dan regresi linear, tujuan dari dilakukannya perbandingan algoritma ini adalah untuk mengetahui nilai akurasi mana yang lebih baik dalam melakukan hal prediksi harga mobil bekas.