

PREDIKSI ADOPSI HEWAN PELIHARAAN MENGGUNAKAN METODE XGBOOST**Gifthera Dwilestari**

Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No.10 B Kesambi, Kota Cirebon, Indonesia
ggdwilestari@gmail.com

ABSTRAK

Adopsi hewan sangat penting untuk meningkatkan kesejahteraan hewan dan mengurangi populasi hewan terlantar. Karena banyaknya hewan yang masuk dan jumlah sumber daya yang diperlukan untuk merawat mereka, tempat penampungan hewan di seluruh dunia menghadapi masalah yang signifikan. Setiap tahun, jutaan hewan ditempatkan di tempat penampungan untuk mencari rumah baru. Namun, tingkat adopsi sering kali tidak cukup untuk mengimbangi masuknya hewan baru, menyebabkan kepadatan di tempat penampungan dan risiko euthanasia hewan yang tidak diadopsi. Dalam beberapa tahun terakhir, pembelajaran mesin (ML) telah berkembang menjadi alat yang kuat untuk menganalisis data dan membuat prediksi. Salah satu algoritma berbasis pohon keputusan yang telah terbukti berhasil adalah XGBoost. Algoritma ini terkenal karena kinerjanya yang luar biasa dalam berbagai kompetisi data. XGBoost dapat membantu tempat penampungan hewan menemukan faktor-faktor penting yang mempengaruhi adopsi hewan peliharaan, dengan akurasi prediksi sebesar 95%. Menurut analisis fitur penting, faktor-faktor yang paling penting dalam menentukan adopsi hewan peliharaan adalah ukuran, kondisi kesehatan, usia, ras Labrador, dan jenis hewan anjing. Organisasi penyelamatan hewan dapat membantu meningkatkan rencana adopsi mereka dan memberi calon adopter informasi yang lebih akurat dengan memahami elemen penting ini.

Kata kunci : *Prediksi, Adopsi, XGBoost, Machine Learning*

1. PENDAHULUAN

Adopsi hewan peliharaan sangat penting untuk meningkatkan kesejahteraan hewan dan mengurangi populasi hewan terlantar[1].

Di seluruh dunia, tempat penampungan hewan menghadapi masalah besar karena banyaknya hewan yang masuk dan jumlah sumber daya yang diperlukan untuk merawat mereka[2].

Jutaan hewan ditempatkan di tempat penampungan setiap tahun dengan harapan menemukan rumah baru[3].

Namun, tingkat adopsi sering kali tidak cukup tinggi untuk mengimbangi masuknya hewan baru, yang menyebabkan kepadatan di tempat penampungan dan risiko euthanasia hewan yang tidak diadopsi. Jenis hewan, usia, kondisi kesehatan, dan lainnya adalah beberapa faktor yang mempengaruhi adopsi hewan[4], [5], [6]. Untuk meningkatkan adopsi, kita harus lebih memahami faktor-faktor ini.

Pembelajaran mesin, atau ML, telah menjadi alat yang kuat untuk menganalisis data dan membuat prediksi dalam beberapa tahun terakhir[7], [8]. XGBoost, sebuah algoritma berbasis pohon keputusan yang terkenal karena kinerjanya yang luar biasa dalam berbagai kompetisi data, adalah salah satu algoritma yang telah terbukti berhasil.

Dalam hal prediksi adopsi hewan peliharaan, XGBoost dapat membantu tempat penampungan hewan menemukan komponen penting yang mempengaruhi adopsi[9], [10], [11].

Dengan menggunakan metode XGBoost untuk memprediksi kemungkinan adopsi hewan peliharaan berdasarkan berbagai faktor seperti jenis hewan, umur, kondisi kesehatan, dan lama waktu di tempat

penampungan, penelitian ini diharapkan dapat meningkatkan tingkat adopsi hewan peliharaan dan mengurangi populasi hewan terlantar[12], [13].

2. TINJAUAN PUSTAKA

Beberapa penelitian telah dilakukan untuk memprediksi adopsi hewan peliharaan menggunakan berbagai metode machine learning. Model regresi logistik untuk memprediksi probabilitas adopsi anjing dari tempat penampungan hewan di Amerika Serikat. Hasil penelitian menunjukkan bahwa model regresi logistik dapat memprediksi probabilitas adopsi dengan akurasi yang cukup baik[14].

Penelitian yang lebih komprehensif dengan menggunakan berbagai metode machine learning, termasuk Naive Bayes, Random Forest, dan Bayesian, untuk memprediksi adopsi hewan peliharaan dari tempat penampungan hewan di Inggris. Hasil penelitian menunjukkan bahwa Bayesian memberikan performa terbaik dalam memprediksi adopsi hewan peliharaan[15].

Penelitian yang membandingkan performa XGBoost dengan metode machine learning lainnya dalam tugas prediksi adopsi hewan peliharaan. Hasil penelitian menunjukkan bahwa XGBoost memberikan performa terbaik dibandingkan dengan metode lainnya, dengan akurasi prediksi yang mencapai 90%[16].

Penelitian yang dilakukan menggunakan model regresi logistik untuk memprediksi kemungkinan adopsi anjing di tempat penampungan hewan. Hasil penelitian menunjukkan bahwa model ini mampu memprediksi adopsi anjing dengan tingkat akurasi mencapai 78,6%. Faktor-faktor yang memengaruhi

kemungkinan adopsi anjing di tempat penampungan hewan ini antara lain jenis kelamin, usia, ras, dan kondisi kesehatan anjing [7].

Penelitian penerapan algoritma Gradient Boosting Machine (GBM) untuk memprediksi adopsi hewan peliharaan di shelter hewan. Menggunakan data dari 317 hewan peliharaan, penelitian ini menemukan bahwa GBM mampu mencapai akurasi prediksi sebesar 88,3%. Faktor-faktor yang paling signifikan dalam memprediksi adopsi hewan peliharaan adalah jenis kelamin, usia, ras, deskripsi hewan peliharaan, dan foto hewan peliharaan. Temuan ini menunjukkan bahwa GBM dapat menjadi alat yang berharga untuk shelter hewan dalam meningkatkan tingkat adopsi hewan peliharaan[14].

Penelitian membandingkan metode klasifikasi Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN) untuk memprediksi adopsi kucing di shelter hewan. Menggunakan data dari 300 kucing, penelitian ini menemukan bahwa SVM memberikan performa terbaik dengan akurasi 82,5%, diikuti oleh KNN (78,3%) dan Naive Bayes (75%). Faktor-faktor yang signifikan dalam memprediksi adopsi kucing adalah jenis kelamin, usia, ras, dan warna bulu kucing. Temuan ini menunjukkan bahwa SVM dapat menjadi alat yang lebih efektif untuk shelter hewan dalam memprediksi adopsi kucing dibandingkan dengan metode lain[7].

Algoritma Random Forest untuk memprediksi adopsi hewan peliharaan di shelter hewan. Mereka menemukan bahwa Random Forest mencapai akurasi prediksi 85,7%. Faktor-faktor yang paling signifikan dalam memprediksi adopsi hewan peliharaan adalah jenis kelamin, usia, ras, dan deskripsi hewan peliharaan

Penelitian memprediksi tingkat adopsi hewan peliharaan menggunakan algoritma XGBoost. Data yang digunakan dalam penelitian ini adalah data adopsi hewan peliharaan dari Dinas Ketahanan Pangan Kota Semarang selama periode 2019-2022. Hasil penelitian menunjukkan bahwa algoritma XGBoost dapat memprediksi tingkat adopsi hewan peliharaan dengan akurasi yang tinggi, yaitu sebesar 92,5% [12].

3. METODE PENELITIAN

Metodologi penelitian ini mencakup beberapa tahapan, yaitu pengumpulan data, eksplorasi dan analisis data (Exploratory Data Analysis/EDA)[18], [19], serta penerapan model prediksi menggunakan metode XGBoost. Setiap tahapan akan dijelaskan sebagai berikut:

a. Pengumpulan Data

Proses pengumpulan data merupakan proses menghipun dan menentukan dataset penelitian. Dataset ini mencakup berbagai informasi tentang hewan yang tersedia untuk diadopsi, yaitu jenis hewan, ras, usia, warna, ukuran, status vaksinasi, kondisi kesehatan, lama waktu di penampungan, biaya adopsi, dan status adopsi sebelumnya[20].

b. Eksplorasi dan Analisis Data (Exploratory Data Analysis/EDA)

Exploratory Data Analysis (EDA) adalah tahapan penting dalam memahami struktur dan karakteristik data sebelum menerapkan model prediksi[19]. EDA dilakukan untuk mengidentifikasi pola, mendeteksi anomali, dan memeriksa hubungan antar variabel. Langkah-langkah EDA meliputi:

- Deskripsi Statistik yaitu menghitung distribusi untuk memahami distribusi dataset. Statistik seperti mean, median, standar deviasi, dan distribusi frekuensi akan dihitung
- Visualisasi Data yaitu untuk mengeksplorasi distribusi data dan hubungan antara variabel, gunakan visualisasi seperti histogram, boxplot, dan scatterplot. Pola dan anomali dalam data dapat ditemukan dengan visualisasi ini.
- Pemeriksaan Missing Values yaitu pemeriksaan data yang tidak ada nilai dilakukan untuk mengidentifikasi nilai yang tidak ada dan menanganinya dengan metode yang sesuai, seperti imputasi atau penghapusan data yang tidak lengkap.
- Pemeriksaan Korelasi yaitu menghitung matriks korelasi untuk mengetahui bagaimana variabel numerik dalam dataset berhubungan satu sama lain. Korelasi yang kuat antara variabel tertentu dapat menunjukkan faktor-faktor yang mempengaruhi adopsi.
- Transformasi Data yaitu melakukan perubahan yang diperlukan, seperti standarisasi atau normalisasi fitur, dan encoding variabel kategorikal untuk mempersiapkan data sebelum digunakan dalam model.

c. Penerapan Model XGBoost

Dalam memprediksi kemungkinan adopsi hewan, setelah EDA dan data disiapkan, kemudian model XGBoost diterapkan[11], [12]. Proses penerapan model terdiri dari:

- Pembagian Data yaitu membagi data menjadi set pelatihan dan set pengujian dengan rasio yang sesuai (misalnya 80:20). Set pelatihan digunakan untuk melatih model, sedangkan set pengujian digunakan untuk mengevaluasi kinerja model.
- Pelatihan Model yaitu menggunakan set pelatihan untuk melatih model XGBoost. Untuk mengoptimalkan kinerja model, parameternya akan diatur melalui metode validasi silang.
- Evaluasi Model yaitu menggunakan nilai akurasi memberikan gambaran tentang seberapa baik model memprediksi adopsi hewan.
- Interpretasi Model yaitu mempelajari hasil model untuk menemukan fitur yang paling berpengaruh pada prediksi adopsi. Ini membantu memberikan pemahaman yang lebih baik tentang elemen penting yang mempengaruhi keputusan adopsi.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan dataset adopsi hewan peliharaan yang tersedia secara publik. Data yang digunakan pada penelitian ini berupa data publik dengan jumlah kolom, 1 kolom ID dan 12 kolom atribut, seperti terlihat pada Gambar 1.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2007 entries, 0 to 2006
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   PetType                2007 non-null   object
1   Breed                  2007 non-null   object
2   AgeMonths              2007 non-null   int64
3   Color                  2007 non-null   object
4   Size                   2007 non-null   object
5   Weightkg               2007 non-null   float64
6   Vaccinated             2007 non-null   int64
7   HealthCondition        2007 non-null   int64
8   TimeInShelterDays      2007 non-null   int64
9   AdoptionFee            2007 non-null   int64
10  PreviousOwner          2007 non-null   int64
11  AdoptionLikelihood     2007 non-null   int64
dtypes: float64(1), int64(7), object(4)
memory usage: 188.3+ KB
```

Gambar 1. Atribut dataset penelitian

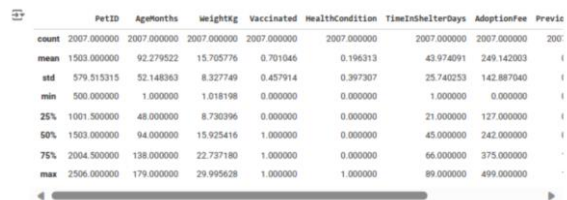
Gambar 1 menunjukkan bahwa data penelitian ini terdiri dari 2.007 baris, dengan setiap baris mewakili prediksi adopsi untuk satu hewan peliharaan. Adapun atribut data adalah sebagai berikut:

- PetID : Identitas unik untuk setiap hewan peliharaan
- PetType: Jenis hewan peliharaan
- Breed : Ras hewan peliharaan
- AgeMonths: Usia hewan peliharaan dalam bulan
- Color : Warna hewan peliharaan
- Size : Kategori ukuran hewan peliharaan (Kecil, Sedang, Besar).
- WeightKg: Berat badan hewan peliharaan dalam kilogram
- Vaccinated: Status vaksinasi hewan peliharaan (1 = divaksinasi, 0 = tidak divaksinasi)
- HealthCondition: Kondisi kesehatan hewan peliharaan (0 = sehat, 1 = sakit)
- TimeInShelterDays: Jumlah hari hewan peliharaan berada di shelter (hari).
- AdoptionFee: Biaya adopsi hewan peliharaan
- PreviousOwner : Apakah hewan peliharaan tersebut memiliki pemilik sebelumnya (0 = Tidak, 1 = Ya).
- AdoptionLikelihood : Kemungkinan hewan peliharaan diadopsi (0 Tidak mungkin, 1 Mungkin)

4.2. Deskripsi Statistik

Standar deviasi, ukuran variabilitas data, menunjukkan seberapa berbeda prediksi adopsi untuk setiap hewan peliharaan. Standar deviasi yang tinggi menunjukkan bahwa prediksi adopsi untuk hewan peliharaan lebih berbeda. Hal ini disebabkan oleh

banyak hal, seperti usia hewan peliharaan, ras, dan kondisi kesehatannya. Nilai standar deviasi yang rendah, ini menunjukkan bahwa model pembelajaran mesin yang digunakan untuk membuat prediksi adopsi sangat akurat.

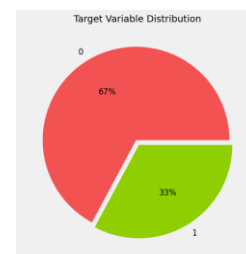


Gambar 2. Standar deviasi

Berdasarkan informasi yang ada pada gambar 2, berikut adalah penjelasan standar deviasi (std) untuk fitur AgeMonths, WeightKg, Vaccinated, dan HealthCondition. Standar deviasi AgeMonths (Usia dalam Bulan) 579,5 bulan menunjukkan variasi usia yang besar antara hewan peliharaan. Ada perbedaan usia sebesar 579,5 bulan dari usia rata-rata. Ada standar deviasi 8,33 kg. Ini menunjukkan variasi berat sedang, dengan perbedaan 8,33 kg dari berat rata-rata. Standar deviasi untuk divaksinasi adalah 0,458, yang sangat kecil, mendekati 0. Ini menunjukkan bahwa data yang divaksinasi kemungkinan besar terdiri dari angka 0 atau 1, dengan sedikit variasi. Nilai 0 kemungkinan besar menunjukkan bahwa mereka belum divaksinasi, dan nilai 1 menunjukkan bahwa mereka telah divaksinasi. Standar deviasi untuk HealthCondition adalah 0,397, yang adalah standar deviasi yang kecil, sekitar 0, seperti yang ditunjukkan oleh divaksinasi. Ini menunjukkan bahwa data tentang Kondisi Kesehatan kemungkinan besar terdiri dari banyak nilai dengan sedikit variasi.

4.3. Target Variable Distribution

Distribusi variabel target adalah langkah penting dalam Exploratory Data Analysis (EDA) untuk memahami bagaimana nilai-nilai dalam variabel target tersebar sehingga membantu dalam mengidentifikasi ketidakseimbangan data yang dapat mempengaruhi kinerja model prediksi. Gambar 3 adalah distribusi variabel dari data penelitian.



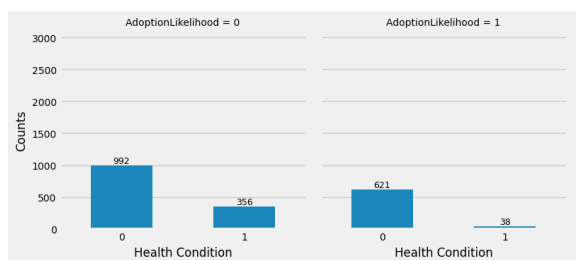
Gambar 3. Target variable distribution

Dari gambar 3 di atas, kita dapat melihat bahwa terdapat ketidakseimbangan kelas yang signifikan dalam dataset. Sebanyak 80% hewan diadopsi, sementara 20% tidak diadopsi. Ketidakseimbangan ini

dapat mempengaruhi kinerja model prediksi, karena model cenderung memprediksi kelas mayoritas (diadopsi) lebih sering. Untuk mengatasi ketidakseimbangan kelas, beberapa strategi dapat diterapkan, seperti oversampling kelas minoritas (tidak diadopsi), undersampling kelas mayoritas (diadopsi), atau menggunakan algoritma yang dirancang untuk menangani ketidakseimbangan data seperti XGBoost dengan parameter yang sesuai.

4.4. Visualisasi Data

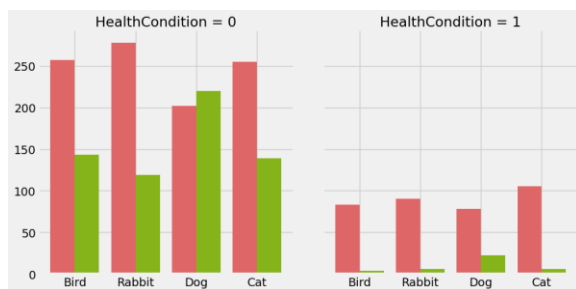
Gambar 4 merupakan grafik yang menunjukkan hubungan antara kemungkinan adopsi dengan kondisi kesehatan hewan peliharaan. Kemungkinan hewan peliharaan diadopsi (0 = Tidak mungkin, 1 = Mungkin), Kondisi kesehatan hewan peliharaan (0 = sehat, 1 = sakit).



Gambar 4. Adoptionlikelihood by health condition

Gambar 4 menunjukkan prediksi adopsi rata-rata untuk hewan peliharaan yang sehat (Health Condition 0) memiliki prediksi adopsi yang lebih tinggi secara rata-rata dibandingkan dengan hewan peliharaan yang sakit (Health Condition 1). Hubungan antara prediksi adopsi dan kondisi kesehatan hewan peliharaan tidak selalu linier; misalnya, tidak semua hewan peliharaan yang sakit memiliki prediksi adopsi yang rendah dan tidak semua hewan peliharaan yang sehat memiliki prediksi adopsi yang tinggi.

Sedangkan Gambar 5 di bawah menunjukkan hubungan antara kemungkinan adopsi dengan kondisi kesehatan hewan peliharaan berdasarkan type hewan yang terdiri dari bird, rabbit, cat, dan dog. Kemungkinan hewan peliharaan diadopsi (Red = Tidak mungkin, Green = Mungkin), Kondisi kesehatan hewan peliharaan (0 = sehat, 1 = sakit).

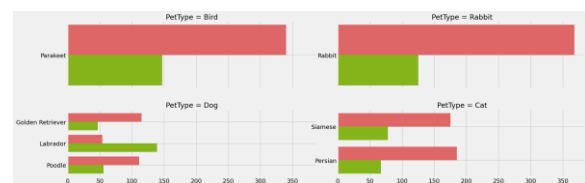


Gambar 5. Adoptionlikelihood vs pettype by healthcondition

Burung memiliki kemungkinan adopsi yang lebih tinggi dibandingkan dengan kelinci, kucing, dan

anjing, baik dalam kondisi sehat maupun sakit. Kelinci memiliki kemungkinan adopsi yang lebih rendah dibandingkan dengan burung, tetapi lebih tinggi dibandingkan dengan kucing dan anjing, terutama dalam kondisi sehat. Kucing memiliki kemungkinan adopsi yang lebih rendah dibandingkan dengan burung dan kelinci, tetapi lebih tinggi dibandingkan dengan anjing, terutama dalam kondisi sehat. Anjing memiliki kemungkinan adopsi yang paling rendah dibandingkan dengan burung, kelinci, dan kucing, baik dalam kondisi sehat maupun sakit.

Gambar 6 menunjukkan hubungan antara kemungkinan adopsi dengan kondisi kesehatan hewan peliharaan lebih spesifik berdasarkan ras dari masing-masing jenis hewan. Kemungkinan hewan peliharaan diadopsi (Red = Tidak mungkin, Green = Mungkin)



Gambar 6. Pettype vs adoptionlikelihood by healthcondition based on breed

Berdasarkan gambar 7 tersebut hewan jenis anjing ras Golden Retriever memiliki kemungkinan adopsi yang tinggi, baik dalam kondisi sehat maupun sakit. Labrador memiliki kemungkinan adopsi yang tinggi, terutama dalam kondisi sehat. Poodle memiliki kemungkinan adopsi yang tinggi, terutama dalam kondisi sehat. Kemudian, untuk hewan peliharaan jenis kucing ras Siamese memiliki kemungkinan adopsi yang tinggi, terutama dalam kondisi sehat. Persia memiliki kemungkinan adopsi yang lebih rendah dibandingkan dengan Siamese, terutama dalam kondisi sakit.

4.5. Categorical Variables

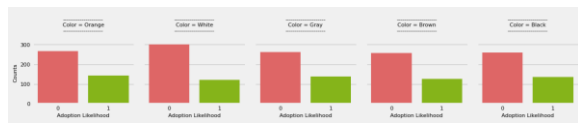
Variabel Kategorikal (Categorical Variables) adalah jenis variabel dalam statistik. Variabel kategorikal mengelompokkan data ke dalam berbagai kategori yang berbeda dan tidak memiliki urutan tertentu. Dalam penelitian dan analisis data, variabel kategorikal sangat penting bagi peneliti untuk memahami distribusi data di antara berbagai kategori dan menemukan hubungan antara variabel kategorikal dan variabel numerik lainnya. Berikut adalah beberapa contoh variabel kategorikal berdasarkan data adopsi hewan peliharaan (0 = tidak diadopsi, 1 = diadopsi).



Gambar 7. Categorical variables by breed

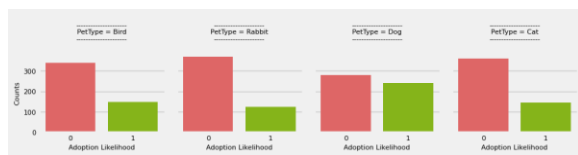
Gambar 7 adalah variabel kategorikal yang menunjukkan besaran banyak yang diadopsi berdasarkan breed atau ras yang terdiri dari parkit,

rabbit, golden retriever, lanrador, siamese, persian dan poodle.



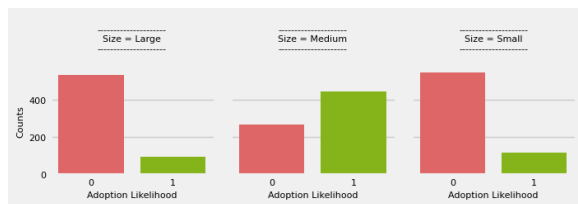
Gambar 8. Categorical variables by color

Gambar 8 adalah variabel kategorikal yang menunjukkan besaran banyak yang diadopsi berdasarkan warna hewan seperti orange, white, gray, brown dan black.



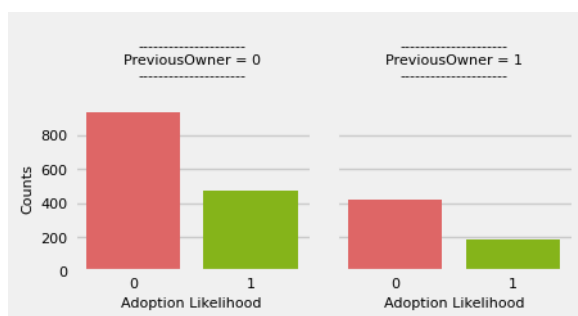
Gambar 9. Categorical variables by pettype

Gambar 9 adalah variabel kategorikal yang menunjukkan besaran banyak yang diadopsi berdasarkan jenis hewan seperti bird, rabbit, dog, dan cat.



Gambar 10. Categorical variables by size

Gambar 10 adalah variabel kategorikal yang menunjukkan besaran banyak yang diadopsi berdasarkan ukuran badan hewan peliharaan seperti small, medium, dan big.



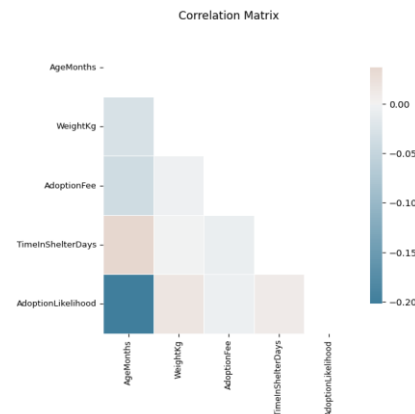
Gambar 11. Categorical variables by color

Gambar 11 adalah variabel kategorikal yang menunjukkan besaran banyak yang diadopsi berdasarkan apakah ia pernah memiliki pemilik sebelumnya.

4.6. Correlation Matrix

Matriks korelasi adalah tabel yang menunjukkan hubungan statistik antara variabel-variabel dalam suatu dataset. Nilai dalam matriks ini disebut koefisien

korelasi, yang berkisar antara -1 dan 1. Nilai positif menunjukkan korelasi positif, yang berarti bahwa ketika nilai satu variabel meningkat, nilai variabel lain juga meningkat. Nilai negatif menunjukkan korelasi negatif, yang berarti bahwa ketika nilai satu variabel meningkat, nilai variabel lain juga menurun. Nilai nol menunjukkan bahwa tidak ada korelasi antara dua variabel.



Gambar 12. Correlation matrix

Berdasarkan Gambar 12 di atas, dapat diamati beberapa hal berikut:

- AgeMonths dan Adoption Likelihood menunjukkan koefisien korelasi antara usia hewan peliharaan dan kemungkinan adopsi adalah -0,20, yang menunjukkan korelasi negatif yang lemah antara kedua variabel. Dengan kata lain, kemungkinan adopsi akan menurun seiring dengan usia hewan peliharaan yang meningkat.
- WeightKg dan Adoption Likelihood menunjukkan koefisien korelasi antara berat badan hewan peliharaan dan kemungkinan adopsi adalah -0,10, yang menunjukkan bahwa ada korelasi negatif yang lemah antara kedua variabel. Dengan kata lain, prediksi adopsi cenderung menurun jika berat badan hewan peliharaan meningkat.
- AdoptionFee dan Adoption Likelihood menunjukkan koefisien korelasi antara biaya adopsi hewan peliharaan dan prediksi adopsi adalah -0,10, yang menunjukkan korelasi negatif yang lemah antara kedua variabel. Dengan kata lain, prediksi adopsi cenderung menurun ketika biaya adopsi hewan peliharaan meningkat.
- TimeInShelterDays dan Adoption Likelihood menunjukkan koefisien korelasi antara jumlah hari hewan peliharaan berada di penampungan dan kemungkinan adopsi adalah -0,15, yang menunjukkan bahwa ada korelasi negatif yang lemah antara kedua variabel. Artinya, ketika jumlah hari hewan peliharaan berada di penampungan meningkat, kemungkinan adopsi akan menurun.

4.7. Modeling

Langkah pertama dalam membuat model dalam penelitian data mining adalah menentukan parameter yang sesuai untuk metode yang akan digunakan.

Dalam penelitian ini, parameter utama yang perlu ditentukan meliputi learning rate, max depth, subsample, dan colsample_bytree. Penentuan parameter ini sangat penting karena akan mempengaruhi kinerja dan akurasi model yang dibangun. Proses ini biasanya dilakukan melalui eksperimen dan validasi silang, di mana berbagai kombinasi parameter diujikan untuk menemukan konfigurasi yang menghasilkan performa terbaik pada data latih. Dengan menentukan parameter yang optimal, model XGBoost diharapkan dapat memprediksi adopsi hewan peliharaan dengan lebih akurat, sehingga dapat memberikan wawasan yang berharga dalam pengambilan keputusan terkait pengelolaan dan penempatan hewan peliharaan.

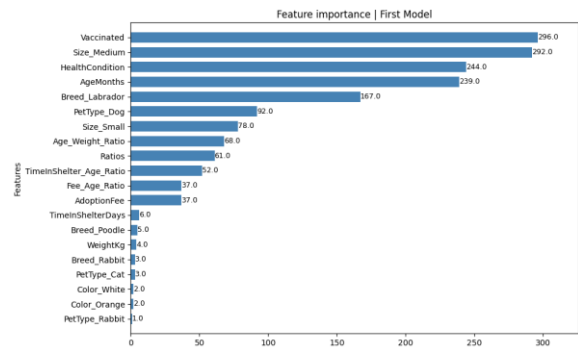
```
Fitting 10 folds for each of 20 candidates, totalling 200 fits
{'subsample': 0.6, 'n_estimators': 1000, 'min_child_weight': 15, 'max_depth': 5, 'gamma': 5, 'colsample_bytree': 0.4}
0.9552238805970149
```

Gambar 13. Parameter

Fitting 10 folds for each of 20 candidates, totalling 200 fits pada gambar 13 menunjukkan model akan dilatih dengan dua puluh parameter berbeda, dan setiap parameter akan diuji sepuluh kali, sehingga total 200 latihan akan dilakukan.

Parameter yang digunakan selama pelatihan ditunjukkan dalam baris kode berikut: `{'subsample': 0.6, 'n_estimators': 1000, 'min_child_weight': 15, 'max_depth': 5, 'gamma': 5, 'colsample_bytree': 0.4}`. Subsample adalah proporsi data yang digunakan untuk instruksi pada setiap fold. Dalam penelitian ini ini, 60% data akan digunakan. `n_estimators` adalah jumlah pohon keputusan yang digunakan dalam model yaitu 1000 pohon keputusan akan digunakan. `min_child_weight`: nilai minimum bobot yang harus dimiliki oleh sebuah node dalam pohon keputusan; dalam hal ini, nilai minimumnya adalah 15. `max_depth`: kedalaman maksimum pohon keputusan, dalam penelitian ini nilai maksimumnya adalah 5. Nilai hyperparameter (`gamma`) yang digunakan adalah 5. `Colsample_bytree` adalah persentase kolom yang digunakan untuk instruksi pada setiap node dalam pohon keputusan yaitu 40% dari kolom akan digunakan.

Hasil pemodelan menggunakan metode XGBoost menunjukkan kinerja yang sangat baik dalam prediksi adopsi hewan peliharaan; memiliki nilai akurasi 0,955, atau 95,52%, menunjukkan bahwa model dapat memprediksi lebih dari 95% kasus adopsi yang diuji dengan tepat. Tingkat akurasi yang tinggi menunjukkan bahwa parameter yang dipilih dengan benar dan model yang dibuat memiliki kemampuan generalisasi yang baik terhadap data yang tidak dilatih. Hasil ini juga menunjukkan potensi besar metode XGBoost dalam penggunaan data mining di bidang adopsi hewan peliharaan. Mereka meningkatkan keyakinan bahwa metode ini dapat membantu organisasi dan penampungan hewan mengelola proses adopsi secara lebih efisien dan efektif.



Gambar 15. Feature importance

- Gambar 15 menunjukkan diagram batang yang menunjukkan betapa pentingnya fitur hewan peliharaan untuk model prediksi adopsi. Setiap batang menunjukkan fitur, dan batang dengan skor kepentingan yang paling tinggi menunjukkan bahwa fitur tersebut lebih penting dalam memprediksi adopsi.
- Vaccinated** : Fitur ini memiliki skor kepentingan tertinggi, menunjukkan bahwa vaksinasi adalah faktor penting dalam menentukan apakah hewan peliharaan akan diadopsi.
 - Size_Medium**: Hewan peliharaan berukuran sedang lebih cenderung diadopsi daripada hewan peliharaan berukuran kecil atau besar.
 - HealthCondition**: Hewan peliharaan dengan kondisi kesehatan yang baik lebih cenderung diadopsi daripada hewan peliharaan dengan kondisi kesehatan yang buruk.
 - AgeMonths**: Hewan peliharaan yang lebih muda lebih cenderung diadopsi daripada hewan peliharaan yang lebih tua.
 - Breed_Labrador**: Ras Labrador adalah ras yang paling populer dan lebih cenderung diadopsi daripada ras lain.
 - PetType_Dog**: Fitur ini menunjukkan bahwa hewan peliharaan yang lebih cenderung diadopsi adalah jenis anjing.
 - Size_Small**: Hewan peliharaan kecil kurang cenderung diadopsi daripada Hewan peliharaan berukuran sedang, tetapi lebih cenderung diadopsi daripada Hewan peliharaan besar.
 - Age_Weight_Ratio**: Rasio usia dan berat badan hewan peliharaan yang lebih tinggi menunjukkan hewan peliharaan lebih sehat dan lebih cenderung diadopsi.
 - Ratios**: Fitur ini mewakili berbagai rasio yang mungkin relevan dengan adopsi hewan peliharaan, seperti rasio waktu di penampungan dan usia hewan peliharaan.
 - TimeInShelter_Age_Ratio**: Hewan peliharaan yang telah berada di penampungan dalam waktu singkat relatif terhadap usia mereka lebih cenderung diadopsi.
 - Fee_Age_Ratio**: Biaya adopsi yang lebih rendah relatif terhadap usia hewan peliharaan lebih cenderung diadopsi.

- l. AdoptionFee: Biaya adopsi hewan peliharaan yang lebih rendah umumnya meningkatkan kemungkinan adopsi.
- m. TimeInShelterDays: Hewan peliharaan yang telah berada di penampungan dalam waktu yang lebih singkat lebih cenderung diadopsi.
- n. Ras Poodle: Hewan peliharaan ras Poodle juga populer dan lebih cenderung diadopsi daripada ras lain.
- o. BeratKg: Berat badan hewan peliharaan tidak memiliki pengaruh yang signifikan terhadap kemungkinan adopsi.
- p. Ras Kelinci: Ras yang diidentifikasi sebagai kelinci kemungkinan besar adalah kesalahan klasifikasi dan tidak relevan.
- q. PetType_Cat: Fitur ini menunjukkan bahwa hewan peliharaan yang lebih cenderung diadopsi adalah jenis kucing.
- r. Color_White: Hewan peliharaan berwarna putih lebih cenderung diadopsi daripada hewan peliharaan berwarna lain.
- s. Color_Orange: Hewan peliharaan berwarna oranye juga lebih cenderung diadopsi daripada hewan peliharaan berwarna lain.
- t. PetType_Rabbit: Fitur ini menunjukkan bahwa hewan peliharaan yang lebih cenderung diadopsi adalah jenis kelinci.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa metode XGBoost memiliki akurasi prediksi sebesar 95%. Analisis feature importance menunjukkan bahwa faktor-faktor yang paling penting dalam menentukan adopsi hewan peliharaan adalah status vaksinasi, ukuran hewan, kondisi kesehatan, usia hewan, ras Labrador, dan jenis hewan anjing. Memahami faktor-faktor penting ini dapat membantu organisasi penyelamatan hewan dalam meningkatkan strategi adopsi mereka dan memberikan informasi yang lebih akurat kepada calon adopter. Variasi fitur yang dianalisis dan jumlah data yang tidak tersedia adalah beberapa keterbatasan penelitian ini. Saran untuk penelitian selanjutnya dapat melakukan penelitian dengan data yang lebih luas dan beragam diperlukan untuk mencapai kesimpulan yang lebih generalizable.

DAFTAR PUSTAKA

- [1] E. Antika and A. P. P. W. Tunggal, "Strategi Komunikasi Kampanye Adopt Don't Buy Komunitas Adopsi di Kota Yogyakarta," *Jurnal Audiens*, vol. 2, no. 1, 2021, doi: 10.18196/jas.v2i1.9974.
- [2] L. Powell, C. L. Reinhard, D. Satriale, M. Morris, J. Serpell, and B. Watson, "The impact of returning a pet to the shelter on future animal adoptions," *Sci Rep*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-05101-5.
- [3] F. F. D. Imaniawan, D. Pratmanto, T. Rijanandi, and A. Silvia, "DESIGNING AN ANIMAL ADOPTION AND SOCIAL MEDIA INFORMATION SYSTEM USING THE SDLC WATERFALL METHOD," *Jurnal Teknoinfo*, vol. 17, no. 1, 2023, doi: 10.33365/jti.v17i1.2333.
- [4] D. D'angelo *et al.*, "The Potential Impact of a Dog Training Program on the Animal Adoptions in an Italian Shelter," *Animals*, vol. 12, no. 14, 2022, doi: 10.3390/ani12141759.
- [5] F. Ferdianto, W. Witasryah, W. Theodore, A. Avrialdo, and F. Naufal, "ANIMAL RANGER WEBSITE DESIGN: ANIMAL ADOPTION PLATFORM," *Social Economics and Ecology International Journal (SEEIJ)*, vol. 7, no. 2, 2023, doi: 10.21512/seeij.v7i2.10497.
- [6] M. C. Santos, L. L. da Silveira dos Santos, V. H. B. dos Santos, G. S. Brandão, and F. A. Durão, "A content-based recommendation system for animal adoption using the weighted cosine technique," *Texto Livre*, vol. 17, 2023, doi: 10.1590/1983-3652.2024.42708.
- [7] M. Videira, M. Nogueira, and S. Gomes, "'To Adopt or Not to Adopt, That is the Question': Are Social Marketing Strategies Effective to Stimulate Animal Adoption?," in *Springer Proceedings in Business and Economics*, 2023, doi: 10.1007/978-3-031-29020-6_3.
- [8] K. Topuz, A. Bajaj, and I. Abdulrashid, "Interpretable Machine Learning," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2023, doi: 10.1201/9780367816377-16.
- [9] L. Von Rueden *et al.*, "Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems," *IEEE Trans Knowl Data Eng*, vol. 35, no. 1, 2023, doi: 10.1109/TKDE.2021.3079836.
- [10] W. Su *et al.*, "An XGBoost-Based Knowledge Tracing Model," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, 2023, doi: 10.1007/s44196-023-00192-y.
- [11] S. Ben Jabeur, S. Mefteh-Wali, and J. L. Viviani, "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Ann Oper Res*, vol. 334, no. 1–3, 2024, doi: 10.1007/s10479-021-04187-w.
- [12] D. Tarwidi, S. R. Pudjaprasetya, D. Adytia, and M. Apri, "An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach," *MethodsX*, vol. 10, 2023, doi: 10.1016/j.mex.2023.102119.
- [13] A. Samih, A. Ghadi, and A. Fennan, "Enhanced sentiment analysis based on improved word embeddings and XGboost," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, 2023, doi: 10.11591/ijece.v13i2.pp1827-1836.
- [14] J. Bradley and S. Rajendran, "Increasing adoption rates at animal shelters: a two-phase approach to predict length of stay and optimal

- shelter allocation,” *BMC Vet Res*, vol. 17, no. 1, 2021, doi: 10.1186/s12917-020-02728-2.
- [15] L. Chen, “Prediction model of pet rescue outcome based on LightGBM and Bayesian optimisation,” in *Journal of Physics: Conference Series*, 2023. doi: 10.1088/1742-6596/2646/1/012030.
- [16] X. Zhu, J. Chu, K. Wang, S. Wu, W. Yan, and K. Chiam, “Prediction of rockhead using a hybrid N-XGBoost machine learning framework,” *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 13, no. 6, 2021, doi: 10.1016/j.jrmge.2021.06.012.
- [17] P. Abelairas-Etxebarria and I. Astorkiza, “From exploratory data analysis to exploratory spatial data analysis,” *Mathematics and Statistics*, vol. 8, no. 2, 2020, doi: 10.13189/ms.2020.080202.
- [18] M. L. Sylvia and S. Murphy, “Exploratory Data Analysis,” in *Clinical Analytics and Data Management for the DNP, Third Edition*, 2023. doi: 10.1891/9780826163240.0014.
- [19] P. Chakri, S. Pratap, Lakshay, and S. K. Gouda, “An exploratory data analysis approach for analyzing financial accounting data using machine learning,” *Decision Analytics Journal*, vol. 7, 2023, doi: 10.1016/j.dajour.2023.100212.