

EVALUASI AUTOML DALAM PREDIKSI TARIF PENERBANGAN: STUDI KASUS ALGORITMA RANDOM FOREST

Arjon Samuel Sitio

Universitas Tjut Nyak Dhien, Sumatera Utara, Indonesia
email: arjonsitio@yahoo.com

Abstract

Airfare prediction is a crucial challenge in the aviation industry, influenced by various dynamic factors such as demand, seasonality, and airline pricing strategies. Traditional machine learning models require extensive manual feature engineering and hyperparameter tuning, which can be time-consuming and complex. To address this issue, this study evaluates the effectiveness of Automated Machine Learning (AutoML) in predicting airfare, with a specific focus on the Random Forest algorithm as a case study. The research methodology involves data preprocessing, feature selection, and the implementation of AutoML to optimize the predictive model. The dataset consists of historical airfare data with multiple influencing variables. AutoML is employed to automate model selection, feature engineering, and hyperparameter tuning, while Random Forest is chosen as the primary algorithm for evaluation. Model performance is assessed using accuracy metrics such as RMSE (Root Mean Square Error) and R^2 (coefficient of determination). The experimental results demonstrate that AutoML significantly enhances prediction accuracy by optimizing hyperparameters and selecting the most relevant features. The Random Forest algorithm, when optimized through AutoML, achieves a higher prediction accuracy compared to its manually tuned counterpart. These findings highlight the potential of AutoML in improving predictive modeling efficiency and accuracy in the airline industry.

Keywords: AutoML, Random Forest, airfare prediction, machine learning, hyperparameter tuning

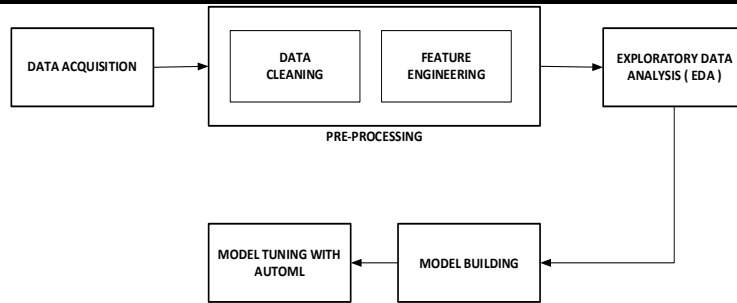
1. PENDAHULUAN

Perkembangan di bidang pariwisata menjadi sebuah daya tarik bagi seorang traveller yang ingin berkunjung dan menikmati wisata di seluruh dunia, dan dapat menjadi dampak positif di sektor penerbangan dan sektor pariwisata. Dengan adanya peran teknologi sangat banyak mempengaruhi segala bidang, dan salah satunya pada penjualan tiket pesawat. Hal tersebut menjadi keuntungan bagi penyedia jasa layanan perjalanan dalam memperjualbelikan tiket pesawat dengan tarif yang murah. Dengan banyaknya maskapai yang bersaing satu sama lain, maka maskapai berlomba untuk menjadi pilihan utama konsumen/pasar, namun untuk mencapai hal tersebut, belum ada strategi maskapai yang dapat memprediksi harga tiket pesawat sesuai dengan kebutuhan pasar. Untuk memenuhi kebutuhan maskapai, maka diperlukan

suatu cara untuk menetapkan harga tiket pesawat sesuai kebutuhan pasar dengan bantuan pengaruh teknologi dan informasi.

Penelitian tentang memprediksi sebuah harga telah banyak dilakukan. Salah satunya penelitian [1]–[3] yang menggunakan algoritma regresi untuk memprediksi pergerakan harga saham menggunakan model *deep learning* dan *fuzzy neural network*. Penelitian [4] untuk memprediksi produksi padi petani dengan model *machine learning* dengan algoritma *decision tree technique* dan data *classification* dengan akurasi 81,6%, Sementara penelitian [5] memprediksi harga bitcoin menggunakan model *machine learning* dengan algoritma *random forest*, *XGBoost*, *Quadratic Discriminant Analysis* dan *Support Vector Machine* dengan dimensi sampel *machine learning techniques*.





Gambar 1. Metode Penelitian

Berdasarkan permasalahan yang telah disebutkan pada paragraf sebelumnya, maka penulis membuat sebuah model *Machine Learning* dengan menggunakan algoritma *Random Forest*, *Logistic Regression* dan *Gradient Boosting Regressor* sehingga pihak maskapai dapat memprediksi penetapan harga tiket pesawat tepat berdasarkan kebutuhan konsumen atau pasar. Parameter yang digunakan dalam membuat sebuah model yaitu jenis

pelayanan, rute penerbangan, durasi penerbangan dan pemberhentian tujuan pesawat tiap maskapai penerbangan. Dari model tersebut diambil algoritma dengan akurasi tertinggi yang akan digunakan untuk memprediksi penetapan tarif tiket pesawat yang akan dikonfigurasi dengan *AutoML*. Dengan model ini memperoleh hasil akurasi minimum di atas 83%.

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302

Gambar 2. Detail Dataset

2. METODE PENELITIAN

Metode penelitian ini dilakukan menggunakan media bantu *Google Colaboratory* untuk membuat sebuah model data dengan algoritma *Random Forest*, *Logistic Regression* dan *Gradient Boosting Regressor*. *Workflow* untuk pembuatan model algoritma yang digunakan dapat dilihat pada Gambar 1. Penelitian diawali dengan pengambilan dataset dari website *Kaggle* yaitu dataset dari *Flight Price Prediction* milik Anshi Gupta; Setelah dataset diunduh, setelah itu dilakukan pada tahap *pre-processing* yang didalamnya terdapat *Data Cleaning* dan *Feature Engineering*; Setelah tahapan *pre-processing* selesai, dilakukan di tahapan *Exploratory Data Analysis (EDA)*; Dan ditahap terakhir yang dilakukan adalah Model

Building tahapan ini terdiri dari pembuatan model untuk memprediksi penetapan harga tarif penerbangan dengan konfigurasi default hingga pembuatan model dengan bantuan *AutoML*.

2.1 Data Acquisition

Penelitian ini, menggunakan dataset yang berasal dari repositori “*Flight Price Prediction*” milik Anshi Gupta di website *Kaggle* [6] agar proses pengambilan data dapat diproses. Dataset ini terdiri 11 kolom yang berisikan data hingga 10683 baris data. Tiap kolom merupakan informasi dari penerbangan seperti pada umumnya yaitu maskapai, tanggal keberangkatan, asal keberangkatan, kedatangan, rute penerbangan, jam keberangkatan, jam kedatangan, durasi penerbangan, dan total pesawat berhenti. Detail dari dataset “*Flight*



Price” yang digunakan dapat dilihat pada Gambar 2.

2.2 Data Cleaning

Tahapan data cleaning berfungsi untuk melakukan proses persiapan menganalisis data yang tidak tepat atau tidak konsisten, dimana data tersebut memiliki pengaruh terhadap model yang akan dibuatkan. Data Cleaning memungkinkan lebih baik dalam pengumpulan data yang representatif untuk meningkatkan tingkat usability dataset yang digunakan dalam segi struktur dan kuliatas data [5]–[8], [9], [10]. Penelitian ini, terdapat nama pada kolom dataset yang digunakan untuk prediksi yang tepat dengan mengubah waktu dan tanggal, kolom dari Date_of_journey akan di ekstrak menjadi 2 kolom disimpan contoh data set saat data cleaning dapat dilihat pada Gambar 3 dan 4.

```
def change_into_datetime(col):  
    df[col]=pd.to_datetime(df[col])  
df.columns  
Index(['Airline',      'Date_of_Journey',  
      'Source', 'Destination', 'Route',  
      'Dep_Time',      'Arrival_Time',  
      'Duration', 'Total_Stops',  
      'Additional_Info', 'Price'],  
      dtype='object')
```

Gambar 3. Mengubah kolom Date_of_Journey

```
df['journey_day']=df['Date_of_Journey']  
.dt.day  
df['journey_month']=df['Date_of_Journey']  
.dt.month  
  
df.drop('Date_of_Journey', axis=1,  
inplace=True)
```

Gambar 4. Menghapus kolom Date_of_journey

2.3 Feature Engineering

Pada Tahapan Feature Engineering merupakan adalah proses mengekstraksi dan memanipulasi karakteristik dari data mentah. representasi numerik dari aspek data mentah berguna untuk membuat lebih banyak fitur atau kolom dari dataset yang telah ada agar terjadi peningkatan kinerja prediksi [8]. Pada penelitian ini, kolom Arrival_time and Dept_time features akan di ekstrak menjadi beberapa kolom baru,

yaitu Dept_time_hour, Dept_time_min dan Arrival_time_hour, arrival_time_min. Syntax untuk Feature Engineering dapat dilihat pada Gambar 5.

```
# function for extracting hour and  
minutes  
def extract_hour(data,col):  
    data[col+'_hour']=data[col].dt.hour  
  
def extract_min(data,col):  
    data[col+'_min']=data[col].dt.minute  
  
def drop_col(data,col):  
    data.drop(col,axis=1,inplace=True)  
  
#call the function  
# Departure time is when a plane  
leaves the gate.  
# Similar to Date_of_Journey we can  
extract values from Dep_Time  
extract_hour(df,'Dep_Time')  
  
#extracting minutes  
extract_min(df,'Dep_Time')  
  
#drop the column  
drop_col(df,'Dep_Time')  
  
#extracting hour  
extract_hour(df,'Arrival_Time')  
  
#extracting min  
extract_min(df,'Arrival_Time')  
  
#drop the column  
drop_col(df,'Arrival_Time')
```

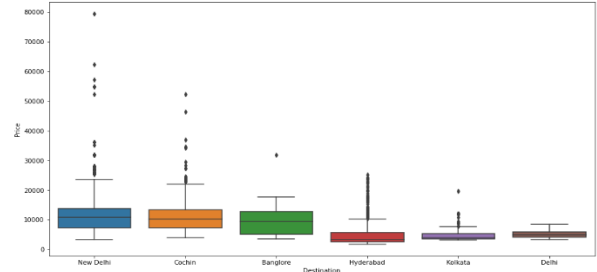
Gambar 5. Syntax Feature Engineering

2.4 Exploratory Data Analysis (EDA)

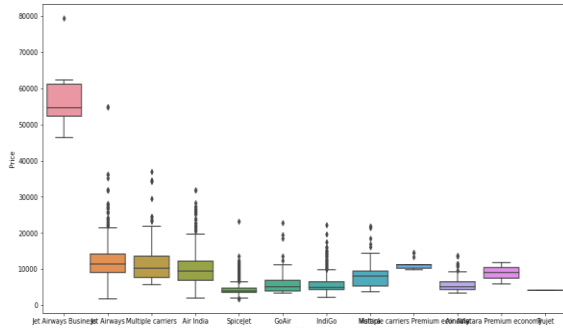
Tahapan Exploratory Data Analysis (EDA) digunakan untuk fungsi dari statistik dan matematik divisualisaikan kedalam bentuk grafik, sehingga mempermudah pemahaman pola data yang ditampilkan. Selain itu, pendekatan tersebut digunakan untuk menganalisa, melihat, dan mencari *outliner* yang ada pada data untuk menarik kesimpulan dari karakteristik utamanya [9]. Pada penelitian ini, EDA dilakukan untuk melihat kolom dengan harga tarif penerbangan berdasarkan kategori data penerbangan. Visualisasi data menggunakan *library seaborn*.



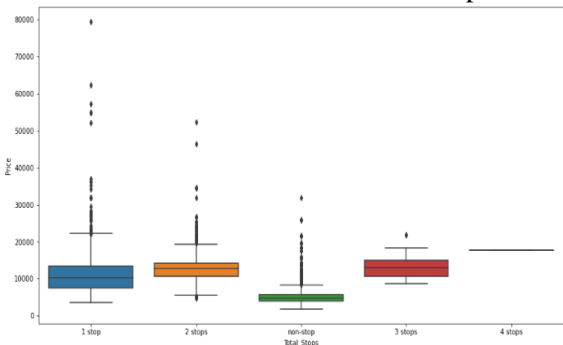
Grafik penerbangan terhadap harga berdasarkan rata-rata maskapai dapat dilihat pada Gambar 6, Grafik penerbangan terhadap harga berdasarkan total_stop dapat dilihat pada Gambar 7, Grafik penerbangan terhadap harga berdasarkan source dapat dilihat pada gambar 8, Grafik penerbangan terhadap harga berdasarkan kategori destinasi dapat dilihat pada gambar 9.



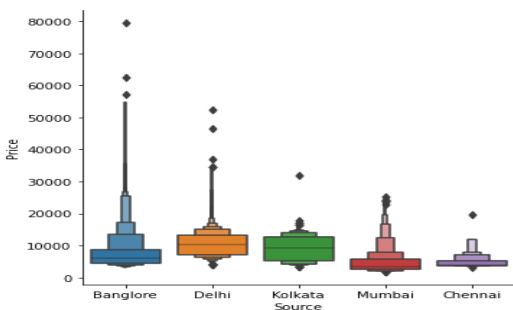
Gambar 5 Grafik destinasi



Gambar 2. Grafik Median Maskapai



Gambar 3 Grafik total_stop



Gambar 4 Grafik Source

2.5 Dataset Splitting

Sebelum pembuatan kinerja model klasifikasi pada algoritma *Machine Learning*, dataset akan dipecah atau dipisahkan dua bagian menjadi data train-testing untuk menghindari *overfitting* [10], [14], [15]. Penelitian ini, konfigurasi yang digunakan untuk *dataset splitting* yang dipakai adalah 7:3, yaitu 70% dari dataset akan menjadi data *training*, sementara 20% dari dataset akan menjadi data *testing*. Proses *dataset splitting* dapat dilihat pada Gambar 10.

```
# splitting the dataset
from sklearn.model_selection import
train_test_split
X_train,X_test,y_train,y_test =
train_test_split(X,y,test_size=0.20,ran
dom_state=123)
```

Gambar 10. Splitting Dataset

2.6 Model Building

Pada penelitian ini, untuk membentuk sebuah model *Machine Learning* yang bisa memprediksi penetapan tarif penerbangan, maka akan dipakai 3 jenis algoritma regresi pada umumnya dipakai dalam pemrosesan data, yaitu *Random Forest*, *Logistic Reggresi* dan *Gradient Boosting*.

2.7 Hypertunning the model

Pada penelitian ini, untuk pengujian nilai secara berulang digunakan untuk proses pelatihan sebuah dataset untuk mendapatkan nilai terbaik dari parameter yang ada. Hal tersebut menjadi poin penting tolak ukur pengujian dengan fitur lengkap dan terpilih [11]. Setelah pembuatan model selesai, maka akan dilakukan *hypertunning* untuk meningkatkan hasil akurasi

model. *Hypertunning* model dilakukan sebagai berikut:

2.7.1. Random Forest Regressor

Random Forest adalah sebuah algoritma *Machine Learning* berbasis *Ensemble Learning* untuk proses klasifikasi dengan membentuk banyak *Decision Tree* untuk mendapatkan proses *training*. Untuk sebuah nilai rata-rata dari prediksi yang dilakukan dari setiap *Tree* akan dilakukan [12]. Pada penelitian ini, algoritma *Random Forest* dibuat menggunakan konfigurasi estimator = 120, max_depth = auto dan max_features = 15. Proses Model *Random Forest* dapat dilihat pada Gambar 11.

```
rf=RandomForestRegressor()  
rf_random=RandomizedSearchCV(estimator=  
rf,param_distributions=random_grid,cv=3  
,verbose=2,n_jobs=-1,)  
  
rf_random.fit(X_train,y_train)  
# best parameter  
rf_random.best_params_
```

Gambar 11. Model *Random Forest*

2.7.2. Logistic Regression

Logistic regression adalah sebuah analisis model linier yang membentuk satu atau sebagian variabel bebas yang menghubungkan satu variabel dengan variabel lain untuk memprediksi dependent[13]. Penelitian ini, algoritma *Logistic Regression* dilakukan dengan konfigurasi default. Proses Model *Logistic Regression* dapat dilihat pada Gambar 12.

$$\pi(x) = \frac{g(x)}{1+g(x)}$$

$$\text{logit}[\pi(x)] = \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] = g(x)$$

```
Model is: LogisticRegression()  
Training score: 0.3356348741954359  
Predictions are: [14571 15237 10844 ...  
4226 7408 10262]
```

```
r2 score is: 0.44250914423182064  
MAE:1872.6485727655593  
MSE:10734864.158165652  
RMSE:3276.4102548621186
```

Gambar 12. model *Logistic Regression*

2.7.3. Gradient Boosting Regressor

Gradient Boosting Regressor sebuah teknik *machine learning* untuk mengatasi masalah dengan klasifikasi, algoritma ini telah terbukti berhasil memprediksi suatu masalah sehingga telah banyak digunakan dalam kompetisi *Kaggle*. Dalam tingkat akurasi, efisiensi, dan interprebilitas model tersebut mencapai kinerja muktahir setiap observasi *training* data [14], [20], [21]. Penelitian ini, algoritma *Gradient Boosting Regressor* dilakukan dengan konfigurasi default. Proses Model *Gradient Boosting Regressor* dapat dilihat pada Gambar 13.

$$\{(x_i, y_i)\}_{i=1}^n$$

$$\frac{\sum \text{Residual}_i}{\sum [\text{Prediksi Sebelumnya}_i \times (1 - \text{Prediksi Sebelumnya}_i)}$$

```
Model is:  
GradientBoostingRegressor()  
Training score: 0.8004872305794601  
Predictions are: [ 5702.86974497  
17826.76035345 12133.74946613 ...  
4547.37535805  
6959.32889786 11428.58468176]  
  
r2 score is: 0.8168061188758857  
MAE:1397.4505612819867  
MSE:3527522.3048542095  
RMSE:1878.1699350309625
```

Gambar 13. Model *Gradient Boosting Regression*

3. HASIL DAN PEMBAHASAN

3.1. Akurasi Model

Penelitian ini menghasilkan sebuah model yang sudah dibentuk, maka komparasi dari model tersebut dapat dilakukan sehingga menghasilkan nilai dari setiap model. Penelitian ini, dikomparasi dari model *Random Forest*, *Logistic Regression*, dan *Gradient Boosting* nilai R2 dan *Real Mean Squared Error* (RMSE) Dapat dilihat pada Tabel 1.

Tabel 1. Komperasi Model



Model	R2 Score	RMSE
1. RF Regressor	83.91%	\$175.9
2. Log Regressor	44.25%	\$327.6
3. GB Regressor	81.68%	\$187.8

Dari tabel diatas, dapat disimpulkan bahwasanya *Random Forest* memiliki nilai R2 tertinggi dan RMSE terendah. Maka *Gradient Boosting* dapat dinyatakan lebih baik dibanding algoritma lainnya untuk dipakai dalam hal memprediksi penetapan tarif penerbangan.

3.2. Feature Importance Model

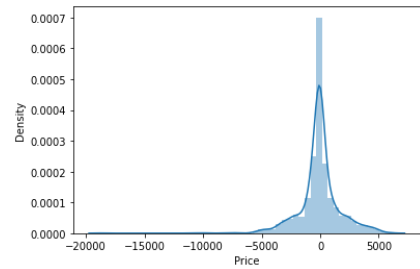
Setelah proses pembuatan model variabel dilakukan, sebagai membantu pihak maskapai sewaktu menetapkan tarif penerbangan sesuai harga pasar, model bisa menjadi acuan menentukan tarif penetapan perbangan yang paling utama. Penelitian ini, bahwasanya dari ketiga algoritma yang dipakai menentukan penetapan tarif penerbangan bahwasanya Route2, Route3, dan Total_Stops adalah bagian fitur terpenting dalam penerbangan untuk penetapan tarif. Visualisasi untuk *Feature Importances* dapat dilihat pada Tabel 2.

Tabel 2. *Feature Importances*

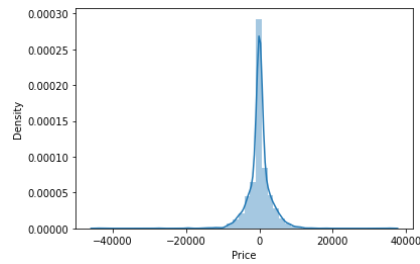
	importance
Route2	2.773248
Route3	2.321282
Total_Stops	2.169140
Route1	2.080196
Arrival_Time_hour	1.874790
dur_hour	1.776536
Arrival_Time_min	1.563630
Delhi	1.547897
Cochin	1.531885
Route4	1.461545
Dep_Time_hour	1.418760
Dep_Time_min	1.217374
dur_min	1.084851

3.3. Prediksi Model

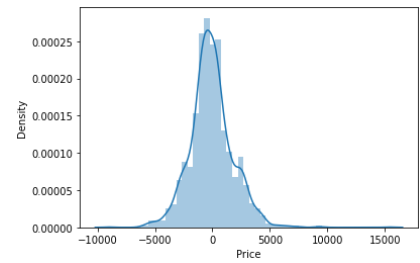
Pada penelitian ini, informasi untuk melihat kinerja apakah model yang telah dibuat dapat memprediksi data valid dengan baik, maka dilakukan uji prediksi terhadap dataset validasi. Ketiga algoritma dipakai dan *scatter plot* akan dipakai untuk melakukan visualisasi prediksi. Visualisasi dapat dilihat pada Gambar 15.



(a)



(b)



(c)

Gambar 15. Scatter Plot Model (a) Random Forest, (b) Logistic Regression, (c) Gradient Boosting

Pada visualisasi diatas, dapat dinyatakan dari ketiga model yang telah dibuatkan mendekati nilai harga asli pada dataset validasi. Maka, dari ketiga model telah dibuat juga dapat dipakai untuk melakukan prediksi penetapan tarif penerbangan.

4. KESIMPULAN

Untuk membantu dalam memprediksi penetapan tarif penerbangan, maka pihak perusahaan maskapai dapat lebih gampang dan waspada untuk menetapkan tarif penerbangan yang sesuai dengan pasar. Dalam penelitian ini, model yang dihasilkan nilai R2 tertinggi dan

RMSE terendah adalah *Random Forest* dengan nilai R2 83.91% dan RMSE \$175.9. Tetapi, dari ketiga model tersebut *Random Forest* mendapat perubahan akurasi sebesar 1,96% menjadi 85,87. Maka dari itu, *Random Forest* dapat dinyatakan lebih baik dari model *Logistic Regression* dan *Gradient Boosting*. Model *Random Forest* yang telah dibuat dapat dipakai untuk memprediksi secara *real-time* menggunakan *Machine Learning*.

5. REFERENSI

- [1] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *J. Financ. Data Sci.*, vol. 4, no. 3, pp. 183–201, 2018, doi: 10.1016/j.jfds.2018.04.003.
- [2] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 6, pp. 1609–1628, 2020, doi: 10.1007/s00521-019-04212-x.
- [3] U. Inyaem, "Construction Model Using Machine Learning Techniques for the Prediction of Rice Produce for Farmers," *2018 3rd IEEE Int. Conf. Image, Vis. Comput. ICIVC 2018*, pp. 870–874, 2018, doi: 10.1109/ICIVC.2018.8492883.
- [4] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, 2020, doi: 10.1016/j.cam.2019.112395.
- [5] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, "Book data grouping in libraries using the k-means clustering method," *J. Phys. Conf. Ser.*, vol. 1230, p. 012074, Jul. 2019, doi: 10.1088/1742-6596/1230/1/012074.
- [6] D. Sitanggang *et al.*, "Diagnosing chicken diseases using fuzzy Tsukamoto web-based expert system," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 505, no. 1, p. 012086, Jul. 2019, doi: 10.1088/1757-899X/505/1/012086.
- [7] A. M. Husein, M. Harahap, S. Aisyah, W. Purba, and A. Muhazir, "The implementation of two stages clustering (k-means clustering and adaptive neuro fuzzy inference system) for prediction of medicine need based on medical data," *J. Phys. Conf. Ser.*, vol. 978, p. 12019, Mar. 2018, doi: 10.1088/1742-6596/978/1/012019.
- [8] W. Purba, S. Tamba, and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, p. 12049, 2018, [Online]. Available: <http://stacks.iop.org/1742-6596/1007/i=1/a=012049>
- [9] E. Saut Parsaoran Tamba, "Penerapan Data Mining Algoritma Apriori Dalam Menentukan Stok Bahan Baku Pada Restoran Nelayan," *JUSIKOM PRIMA (Jurnal Sist. Inf. Ilmu Komput. Prima)*, vol. 5, no. 2, pp. 97–102, 2022.
- [10] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.
- [11] C. M. Sitorus, A. Rizal, and M. Jajuli, "Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 254–265, 2020, doi: 10.28932/jutisi.v6i2.2672.
- [12] R. S. Oktavian and S. Budi, "Analisis Dataset Google Playstore Menggunakan Metode Exploratory Data Analysis," *J. Strateg. Maranatha*, vol. 2, no. 2, pp.



- 636–649, 2020.
- [13] A. Nurhopipah and U. Hasanah, “Dataset Splitting Techniques Comparison For Face Classification on CCTV Images,” vol. 14, no. 4, pp. 341–352, 2020.
- [14] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, “Book data grouping in libraries using the k-means clustering method,” *J. Phys. Conf. Ser.*, vol. 1230, no. 1, p. 012074, Jul. 2019, doi: 10.1088/1742-6596/1230/1/012074.
- [15] V. M. Mulia Siregar and H. Sugara, “Implementation of artificial neural network to assesment the lecturer’s performance,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, p. 012112, Oct. 2018, doi: 10.1088/1757-899X/420/1/012112.
- [16] A. Ramdan, U. Siliwangi, N. Widyasono, U. Siliwangi, H. Mubarok, and U. Siliwangi, “Prediksi Jaringan TOR dan VPN menggunakan Algoritma K-Nearest Neighbour pada Trafik Darknet,” vol. 05, no. 01, pp. 21–35, 2022.
- [17] N. G. Ramadhan, F. D. Adhinata, A. Jala, T. Segara, and D. Putra, “Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression,” vol. 9, no. 2, pp. 251–256, 2022, doi: 10.30865/jurikom.v9i2.3979.
- [18] M. Nawawi, “Klasifikasi Tingkat Popularitas Siswa Berdasarkan Aktifitas Komunikasi Siswa Menggunakan Smartphone dengan Teknik Logistic Regression,” vol. 4, no. 1, pp. 978–979, 2018.
- [19] E. N. Fauziyah and S. R. Nudin, “Sistem Pendukung Keputusan Penentuan Jurusan di SMKN 1 Pungging Menggunakan Gradient Boosting Tree,” vol. 3, pp. 42–50, 2021.
- [20] S. P. Tamba, A. W. Tan, Y. Gunawan, and ..., “Penerapan Data Mining Untuk Pembuatan Paket Promosi Penjualan Menggunakan Kombinasi Fp-Tree Dan Tid-List,” ... (*Teknik Inf. dan ...*, vol. 4, 2021, [Online]. Available: <http://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/309%0Ahttp://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/download/309/239>
- [21] S. P. Tamba and E. -, “Prediksi Penyakit Gagal Jantung Dengan Menggunakan Random Forest,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 176–181, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2445.

