

Analisis Perbandingan *Decision Tree*, *Support Vector Machine*, dan *Xgboost* dalam Mengklasifikasi *Review Hotel Trip Advisor*

Hansen Christanto¹⁾, Julfikar Rahmad²⁾, Stiven Hamonangan Sinurat³⁾, Daniel Ryan Hamonangan Sitompul⁴⁾, Andreas Sitomorang⁵⁾, Dennis Jusuf Ziegel⁶⁾, Evta Indra^{*7)}

¹⁾²⁾³⁾⁴⁾⁵⁾⁶⁾⁷⁾ Sistem Informasi, Sains dan Teknologi, Universitas Prima Indonesia

Correspondence author: evtaindra@unprimdn.ac.id, Medan, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i1.1429>

Abstrak

Jaringan media sosial pada saat ini terus berkembang dan berdampak pada industri perhotelan. Pelanggan dan *traveler* telah memposting hasil *review* secara online untuk menunjukkan tingkat kepuasan mereka terhadap hotel dan berbagi pengalaman terkait hotel yang dikunjungi dengan pelanggan lain yang ada di seluruh belahan dunia. Situs web yang bergerak dalam pariwisata dan perhotelan berkembang pesat secara online seperti *Trip advisor*. *Trip advisor* merupakan *platform* penyedia layanan perjalanan dan pemesanan hotel. Penelitian menggunakan teknik analisis sentimen untuk mengkategorikan opini pengguna yang bernilai negatif maupun positif dengan bantuan kecerdasan buatan yaitu *Machine Learning*. Penelitian ini menguji tiga algoritma *Machine Learning*, yaitu *Decision Tree Classifier*, *Support Vector Machine* (SVM) dan *Xgboost Classifier*, dalam melakukan analisis sentimen terhadap *review* hotel di *platform Trip advisor*. Hasilnya menunjukkan bahwa *Xgboost* memiliki tingkat keakuratan (*accuracy*) yang paling tinggi, mencapai 99%, dibandingkan dengan *Decision Tree* (97%) dan *Support Vector Machine* (98%). Dengan demikian, *Xgboost* dianggap sebagai algoritma terbaik untuk melakukan analisis sentimen pada *review* hotel di *Trip advisor*.

Kata Kunci: *Trip advisor*, *Machine Learning*, *Decision Tree*, *Support Vector Machine*, *Xgboost*

Abstract

Social media networks are currently growing and impacting the hospitality industry. Customers and travelers have posted online review results to show their level of satisfaction with hotels and share experiences regarding hotels visited with other customers around the world. Websites engaged in tourism and hospitality are thriving online such as Trip advisor. Trip Advisor is a platform that provides travel and hotel booking services. This research uses sentiment analysis techniques to categorize user opinions that are negative or positive with the help of artificial intelligence, namely Machine Learning. This study tested three Machine Learning algorithms, namely the Decision Tree Classifier, Support Vector Machine (SVM) and Xgboost Classifier, in conducting sentiment analysis of hotel reviews on the Trip advisor platform. The results show that Xgboost has the highest level of accuracy, reaching 99%, compared to the Decision Tree (97%) and Support Vector Machine (98%). Thus, Xgboost is considered the best algorithm for performing sentiment analysis on hotel reviews on Trip advisor.

Keywords: *Tripadvisor*, *Machine Learning*, *Decision Tree*, *Support Vector Machine*, *Xgboost*

PENDAHULUAN

Jaringan media sosial pada saat ini terus berkembang dan berdampak pada industri perhotelan. Pelanggan dan *traveler* telah memposting hasil *review* secara online untuk menunjukkan tingkat kepuasan mereka terhadap hotel dan berbagi pengalaman terkait hotel yang dikunjungi dengan pelanggan lain yang ada di seluruh belahan dunia. Situs web yang

bergerak dalam pariwisata dan perhotelan berkembang pesat secara online seperti *Trip advisor*. *Trip advisor* merupakan *platform* penyedia layanan perjalanan dan pemesanan hotel. Data yang diperoleh dalam penelitian ini berisikan *dataset* hotel yang ada di kota *Mussoorie*, data yang digunakan berasal dari *platform* perjalanan online asal amerika yaitu: *Trip advisor* yang ada di *Kaggle*. Penelitian menggunakan teknik analisis sentimen untuk mengkategorikan opini pengguna yang bernilai negatif maupun positif dengan bantuan kecerdasan buatan yaitu *Machine Learning*.

Banyak sekali informasi yang ditampilkan di website *Trip advisor*, salah satunya adalah klasifikasi hotel berdasarkan *review* pelanggan. Ada beberapa klasifikasi hotel yang memiliki rating tertinggi hingga terendah. *Review* hotel sangat perlu untuk para *traveler*, dengan adanya *review* tersebut maka terciptanya sebuah rating sehingga para *traveler* dapat memutuskan hotel mana yang sesuai dengan keinginan mereka untuk menginap. Permasalahan yang dapat ditimbulkan ketika pelanggan memberikan hasil *review* mereka terhadap hotel yang pernah dikunjungi bisa berisi komentar yang positif atau negatif, hal tersebut dapat menyulitkan dan membuat bingung para pengunjung yang ingin memesan hotel melalui *platform Trip advisor*. Oleh karena itu perlu dilakukan analisis sentimen terhadap hasil *review* para pelanggan agar pengunjung tidak sulit dalam menentukan hotel mana yang bagus atau tidak sesuai *review* oleh pengunjung sebelumnya.

Telah banyak penelien-penelitian sebelumnya yang sudah melakukan analisis sentimen *review* hotel, sebagai contoh “*Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier*” dan “*Sentiment Analysis on Chinese Hotel Reviews with Doc2Vec and Classifiers*”. Penelitian yang dilakukan menggunakan data hotel yang ada di *platform Trip advisor* untuk menganalisis sentimen sebuah *review* hotel yang dilakukan oleh pelanggan dengan menggunakan metode yang berbeda. Untuk penelitian pertama menggunakan metode *Multinomial Naïve Bayes Classifier* dengan hasil *preprocessing+BoW (bag of words)* sebesar 90.927%, *preprocessing+feature selection-1* sebesar 91,414% dan *preprocessing+featur selection-2* sebesar 91.233%. Sedangkan penelitian kedua menggunakan metode *Doc2vec* dan pengklasifikasian algoritma dan mendapatkan hasil akhir algoritma *SVM (Support Vector Machine)* sebesar 79.5%, *Algoritma Logistic Regression* sebesar 78.83% dan *Naïve Bayes* sebesar 73.42%. Hasil perbandingan penelitian pertama dan kedua dapat dilihat pada tabel dibawah ini:

Tabel 1. Hasil Perbandingan Penelitian Pertama dan Kedua

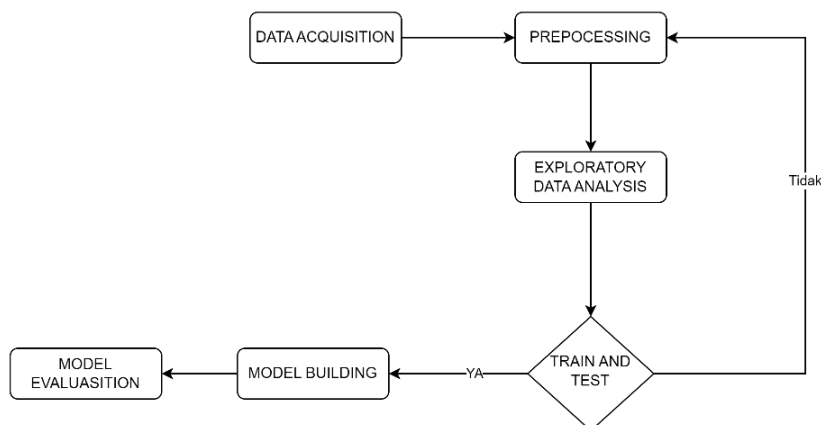
Penelitian I	Hasil penelitian	Penelitian II	Hasil penelitian
preprocessing+BoW	90.927%	Algoritma SVM	79.5%
preprocessing+feature selection -1	91,414%	Algoritma Logistic Reegression	78.83%
preprocessing+featur selection-2	91.233%	Naïve Bayes	73.42%

Pada perbandingan penelitian sebelumnya didapatkan hasil akurasi tertinggi dalam review hotel sebesar 91,414% oleh karena itu berlu dilakukan pengembangan dalam melakukan analisis *review* hotel dengan menggunakan algoritma yang berbeda agar mendapatkan hasil akurasi lebih baik dari penelitian sebelumnya. Berdasarkan latar belakang yang telah dipaparkan, peneliti tertarik untuk melakukan penelitian dengan judul “Analisis Perbandingan *Decision Tree*, *Support Vector Machine*, dan *Xgboost* dalam Mengklasifikasi *Review* Hotel pada *Trip Advisor*”.

METODE

Pada era teknologi seperti sekarang ini. Hotel merupakan sarana yang paling banyak dicari pelancong atau *traveler*. Karena hal tersebut, banyak sekali hotel yang berlomba-lomba dalam mempromosikan hotel mereka di *platform* online seperti *Trip advisor*. Oleh karena itu dalam penelitian ini peneliti ingin mengklasifikasikan setiap hasil *review* pelanggan hotel sehingga dapat diketahui apakah komentar dari pelanggan mengarah ke hasil positif atau negative. Dalam hal ini peneliti menggunakan metode pengolahan bahasa alami yaitu *Natural Language Processing* (NLP) khususnya *Extreme Gradient Boosting package* (Xgboost). Xgboost merupakan model linier yang efisien dalam memecahkan model linier dan mendukung sejumlah proses objektif, seperti regresi, klasifikasi, dan peringkat. *Support Vector Machine* (SVM) adalah algoritma *Machine Learning* yang termasuk dalam model *Supervised Learning* atau pembelajaran terawasi terkait analisis data dan pengenalan pola. *Decision Tree* adalah teknik penambangan data yang biasa digunakan untuk membangun sistem klasifikasi berdasarkan beberapa kovariat atau untuk mengembangkan algoritma prediksi untuk variabel target.

Agar penelitian ini dapat berjalan dengan baik dan selesai dengan tepat waktu maka dibuat prosedur kerja. Adapun prosedur kerja penelitian ini adalah sebagai berikut:



Gambar 1. Flowchat Prosedur Kerja Penelitian

Tahap *Data Acquisition* adalah proses pengambilan data dari sumber yang valid sebelum data tersebut diolah dan digunakan untuk membuat model *Machine Learning*. Akhir dari proses ini adalah memanggil *Data Frame* pada Notebook. Pada penelitian ini data yang dipakai berasal dari *Kaggle* dan dari *Trip advisor*. *Dataset* berisikan 30391 baris dan 3 kolom. Berikut sample *dataset* dari review hotel *Trip Advisor*.

	Review	Date of stay	Rating
0	Most beautiful stay at Mussoorie...such a gorg...	Date of stay: August 2021	
1	NaN	NaN	NaN
2	The Savoy is one of those rare hotels that not...	Date of stay: October 2020	
3	NaN	NaN	NaN
4	Def Best stay in Uttarakand ! Amazing nature. ...	Date of stay: November 2021	
...
30386	It was a wonderful experience in Mussoorie. Gr...	Date of stay: May 2009	
30387	NaN	NaN	NaN
30388	Sterling Resort at the first look doesnt seem ...	Date of stay: April 2009	
30389	NaN	NaN	NaN
30390	Stay at Hotel Sterling Resorts was a good expe...	Date of stay: June 2008	

30391 rows x 3 columns

Gambar 2. Dataset Review Hotel Trip Advisor

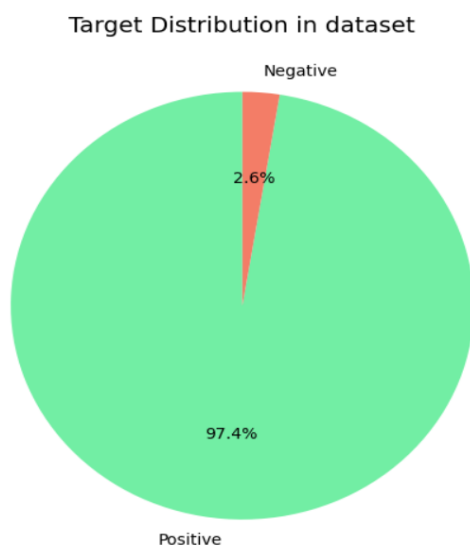
Fase *preprocessing* adalah proses menyiapkan data yang cocok untuk membangun dan melatih model *Machine Learning*. Prosedur ini juga berarti mengonversi data ke format yang lebih mudah dibaca oleh model, *preprocessing* juga membantu meningkatkan kualitas data. Berikut sample gambar dari Fase *preprocessing*.

	Review	Date of stay	Rating
0	Most beautiful stay at Mussoorie...such a gorg...	Date of stay: August 2021	
2	The Savoy is one of those rare hotels that not...	Date of stay: October 2020	
4	Def Best stay in Uttarakand ! Amazing nature. ...	Date of stay: November 2021	
6	Very very friendly and efficient staff ... loved...	Date of stay: November 2021	
8	Exceptional property. Just great experience. V...	Date of stay: October 2021	
...
30382	I stayed in this hotel with my 5 family member...	Date of stay: May 2008	
30384	the best histroacally place in india	Date of stay: April 2007	
30386	It was a wonderful experience in Mussoorie. Gr...	Date of stay: May 2009	
30388	Sterling Resort at the first look doesnt seem ...	Date of stay: April 2009	
30390	Stay at Hotel Sterling Resorts was a good expe...	Date of stay: June 2008	

12796 rows x 3 columns

Gambar 3. Data Preprocessing

Exploratory data analysis (EDA) adalah proses eksplorasi data, yang tujuannya untuk memahami isi dan komponen yang membentuk data. Dalam penelitian ini, EDA dilakukan pada *dataset* untuk melihat isi data, korelasi data, distribusi data. Visualisasi *distribution plot* digunakan untuk melihat deskripsi *review* pengunjung hotel apakah negatif atau positif yang ditampilkan secara visual sehingga dapat lebih dipahami cara membaca data yang ada didalam dataset. Berikut ini sample *Exploratory data analysis review* hotel *Trip advisor*.



Gambar 4. Exploratory data analysis review komentar positif dan negative

Pada proses *Train and Test*, sekumpulan data diproses menggunakan algoritma *Extreme Gradient Boosting package (Xgboost)*, *Support Vector Machine (SVM)* dan *Decision Tree* untuk melakukan analisis *review* pengunjung hotel menggunakan bahasa pemrograman Python. Setelah mengolah data dan memperoleh hasil yang tepat, ditentukan apakah hasil yang diperoleh memuaskan dari segi kinerja. Jika data yang diolah tidak memberikan hasil yang baik setelah melakukan analisis sentimen, data akan kembali ke tahap *data Cleaning* untuk memastikan bahwa hasilnya lebih baik dari penelitian sebelumnya. Berikut sample gambar proses *Train and Test* seperti dilihat pada gambar dibawah ini.

```
# Initializing TfidfVectorizer
tfv = TfidfVectorizer(min_df=3, max_features=None, decode_error = "replace",
                    preprocessor = preprocess_data,
                    strip_accents='unicode', analyzer='word', token_pattern=r'\w{1,}',
                    ngram_range=(1, 3), use_idf=1, smooth_idf=1, sublinear_tf=1,
                    stop_words = 'english')

# Fitting Tfidf vectorizer to training dataset
tfv.fit(list(xtrain))

# Transforming the training and validation dataset
xtrain_tfv = tfv.transform(xtrain)
xval_tfv = tfv.transform(xval)
```

Gambar 5. Proses *Train and Test* pada dataset

Pada penelitian ini akan dibuat model klasifikasi *review* hotel yang dilakukan pengunjung berdasarkan pendapat (opini). Dalam hal ini algoritma yang digunakan adalah *Extreme Gradient Boosting package (Xgboost)*, *Support Vector Machine (SVM)* dan *Decision Tree*. Berikut ini gambar model building saat melakukan pemodelan.

```
#import library DecisionTreeClassifier
from sklearn.tree import DecisionTreeClassifier
DT= DecisionTreeClassifier()
DT.fit(xtrain_tfv,ytrain)

DecisionTreeClassifier()
```

Gambar 6. Model Building

Model Evaluation merupakan hasil akhir dari pengolahan model *Machine Learning* dengan menggunakan *Extreme Gradient Boosting package* (Xgbost), *Support Vector Machine* (SVM) dan *Decision Tree* sehingga hasil dari algoritma tersebut dapat dibandingkan dalam menentukan hotel mana yang terbaik untuk tempat menginap berdasarkan hasil review yang diberikan oleh pengunjung sebelumnya. Berikut sample gambar *Model Evaluation* dapat dilihat.

```
# Transforming test dataset using the same tfidf vectorizer we fit on training dataset
xtest_tfv = tfv.transform(xtest)

#XGBOOST Classifier
# Making prediction on testing dataset
predictions = xgb.predict_proba(xtest_tfv)
print ("logloss: %0.3f " % log_loss(ytest, predictions))
ypred_xg = xgb.predict(xtest_tfv)
print(classification_report(ytest, ypred_xg))
```

logloss: 0.048				
	precision	recall	f1-score	support
0	0.90	0.44	0.59	43
1	0.99	1.00	0.99	1616
accuracy			0.98	1659
macro avg	0.95	0.72	0.79	1659
weighted avg	0.98	0.98	0.98	1659

Gambar 7. Model Evaluation

HASIL DAN PEMBAHASAN

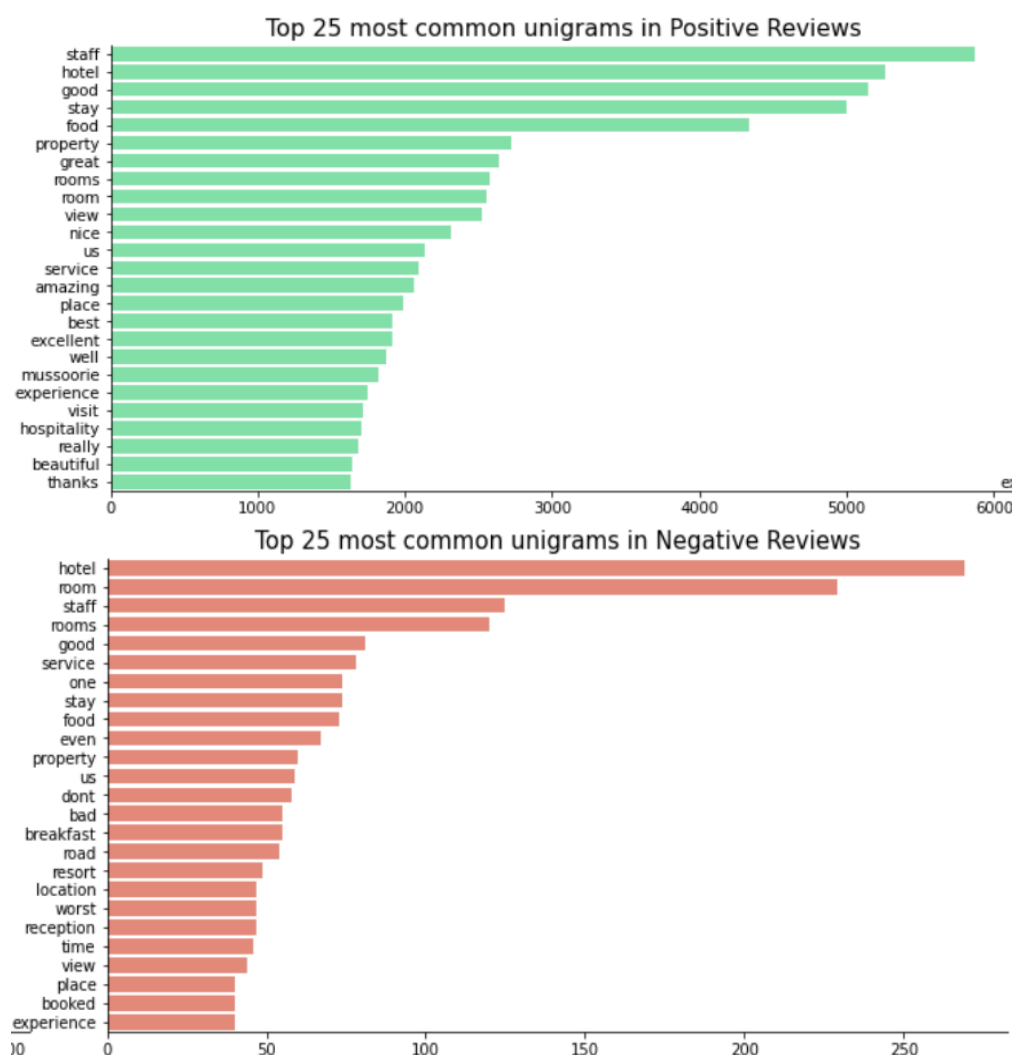
Tahap *cleaning* data atau penghapusan data bertujuan membersihkan data yang akan dipakai untuk membuat model serta dilakukan pengolahan sehingga menghasilkan hasil output yang baik. Dalam penelitian ini *cleaning* data yang dilakukan adalah menghapus data yang tidak bernilai atau disebut dengan *NaN* dan menghapus data yang duplikat. Seperti dapat dilihat pada gambar berikut.

```
#Dropping all the rows with null values
df.dropna(inplace=True)

#Dropping duplicate rows
df.drop_duplicates(inplace=True)
```

Gambar 8. Cleaning Data NAN dan Duplikat

Berdasarkan target distribution pada *dataset* ada sebanyak 97,4% *traveler* yang memberikan komentar positif sedangkan *traveler* dan pengunjung yang memberikan komentar negatif ada sebanyak 2,6% terhadap hotel yang ada pada platform *Trip advisor*. Pada *Exploratori* data *plotting* ini banyak pengunjung yang memberikan penilaian terhadap hotel yang mereka sewa, ada 25 hal yang paling sering dikomentari oleh *traveler* dan pengunjung hotel pada saat menginap, pengunjung hotel yang menggunakan layanan platform *Trip advisor* memberikan 6000 komentar positif terhadap staff hotel, 5300 pengunjung memberikan komentar positif terhadap hotel, sedangkan pengunjung yang memberikan komentar negatif terhadap hotel sebanyak 300 komentar dan yang berkomentar negatif terhadap ruangan ada sebanyak 230 komentar, dapat dilihat pada gambar berikut ini.



Gambar 9. *Top 25 Most Common Unigrams in Negative and Positive Reviews*

Tahap *Train and Test* merupakan pengujian model dengan menggunakan variabel yang telah dibuat sebelumnya. Pada penelitian ini yang menjadi variabel X adalah *review* sedangkan untuk variabel Y adalah sentimen, seperti yang dapat dilihat pada gambar dibawah ini.

```
# Splitting the dataset into training and testing dataset
xtrain, xtest, ytrain, ytest = train_test_split(df['Review'],df['Sentiment'],
                                             stratify=df['Sentiment'],
                                             random_state=42,
                                             test_size=0.2, shuffle=True)

# Splitting training dataset into training and validation dataset
xtrain, xval, ytrain, yval = train_test_split(xtrain, ytrain,
                                             stratify=ytrain,
                                             random_state=42,
                                             test_size=0.1, shuffle=True)
```

Gambar 10. Proses *Train and Test model*

Tahapan *model building* merupakan proses memasukkan *library* dari algoritma yang digunakan untuk memproses model agar dapat menemukan hasil akurasi dari model yang dibuat pada penelitian ini. Algoritma yang digunakan berupa *Xgboost*, *Decision Tree*, dan *Support Vector Machine* (SVM), seperti dilikat pada gambar dibawah ini.

```
# Initializing XGBClassifier
xgb = XGBClassifier(learning_rate=0.02, n_estimators=600, max_depth=4,
                  subsample= 1.0, min_child_weight=1, gamma= 0.5,
                  colsample_bytree= 0.8, random_state=42)

# Fitting xgb on training dataset
xgb.fit(xtrain_tfv, ytrain)

# Making predictions on validation dataset
predictions = xgb.predict_proba(xval_tfv)
print ("logloss: %0.3f " % log_loss(yval, predictions))

ypred_xg = xgb.predict(xval_tfv)
print(classification_report(yval, ypred_xg))
```

Gambar 11. Algoritma Xgboost

```
#import library DecisionTreeClassifier
from sklearn.tree import DecisionTreeClassifier
DT= DecisionTreeClassifier()
DT.fit(xtrain_tfv,ytrain)

DecisionTreeClassifier()

# Making predictions on validation dataset
predictions = DT.predict_proba(xval_tfv)
print ("logloss: %0.3f " % log_loss(yval, predictions))

ypred_DT = DT.predict(xval_tfv)
print(classification_report(yval, ypred_DT))
```

Gambar 12. Algoritma *Decision Tree*

```
# making library algoritma
from sklearn import svm
from sklearn.metrics import accuracy_score

svmc= svm.SVC()
svmc.fit(xtrain_tfv,ytrain)

SVC()

# Making predictions on validation dataset
predictions = svmc.predict(xval_tfv)
print ("logloss: %0.3f " % log_loss(yval, predictions))

ypred_svmc = svmc.predict(xval_tfv)
print(classification_report(yval, ypred_svmc))
```

Gambar 13. Algoritma *Support Vector Machine (SVM)*

Hasil dari *model evaluation* dengan menggunakan algoritma Xgboost didapatkan hasil *precision* yang bernilai negative sebesar 100%, sedangkan hasil *precision* yang bernilai positive sebesar 99%, untuk nilai *recall negative* sebanyak 53% sedangkan yang bernilai *positive* sebesar 100%, untuk nilai *F1-score negative* sebanyak 69% serta nilai *F1-score positive* sebanyak 99%, dan *accuracy* yang didapat dengan menggunakan algoritma Xgboost sebesar 99%. Seperti dapat dilihat pada gambar berikut.

```

logloss: 0.04
      precision    recall  f1-score   support

     0         1.00      0.53      0.69         17
     1         0.99      1.00      0.99        647

 accuracy          0.99          664
 macro avg          0.99      0.76      0.84          664
 weighted avg          0.99      0.99      0.99          664
    
```

Gambar 14. Hasil Pengolahan Algoritma Xgboost

Hasil dari model evaluation dengan menggunakan algoritma *Decision Tree* didapatkan hasil *precision negative* sebesar 50%, sedangkan untuk *precision positive* sebesar 99%, hasil *recall negative* 53%, hasil *recall positive* 99%, hasil *F1-score negative* sebesar 51%, hasil *F1-score positive* 99% dan untuk *accuracy* dari algoritma *Decision Tree* sebesar 97%. Seperti dapat dilihat pada gambar berikut.

```

logloss: 0.884
      precision    recall  f1-score   support

     0         0.50      0.53      0.51         17
     1         0.99      0.99      0.99        647

 accuracy          0.97          664
 macro avg          0.74      0.76      0.75          664
 weighted avg          0.98      0.97      0.97          664
    
```

Gambar 15. Hasil Pengolahan Algoritma *Decision Tree*

Hasil dari model evaluation dengan menggunakan algoritma *Support Vector Machine* (SVM) didapatkan hasil *precision negative* sebesar 100%, sedangkan untuk *precision positive* sebesar 98%, hasil *recall negative* 6%, hasil *recall positive* 100%, hasil *F1-score negative* sebesar 11%, hasil *F1-score positive* 99% dan untuk *accuracy* dari algoritma *Support Vector Machine* (SVM) sebesar 98%. Seperti dapat dilihat pada gambar berikut.

logloss: 0.832					
	precision	recall	f1-score	support	
0	1.00	0.06	0.11	17	
1	0.98	1.00	0.99	647	
accuracy			0.98	664	
macro avg	0.99	0.53	0.55	664	
weighted avg	0.98	0.98	0.97	664	

Gambar 16. Hasil Pengolahan Algoritma *Support Vector Machine* (SVM)

KESIMPULAN DAN REKOMENDASI

Trip advisor adalah salah satu *platform* penyedia jasa layanan perjalanan yang sekarang ini banyak dikunjungi oleh orang di seluruh belahan dunia. Salah satu layanan dari *platform* tersebut yakni pemesanan hotel yang dimana para *traveler* selalu memberikan *review* mereka setelah mereka menginap. Banyak dari *traveler* yang akan menginap mengalami kesulitan dalam menentukan pilihan hotel. Hal tersebut terjadi dikarenakan terlalu banyak *review* dari pengunjung yang pernah menginap di hotel tersebut. Berdasarkan penelitian yang telah dilakukan dengan menggunakan tiga algoritma yaitu *Xgboost Classifier*, *Decesiontree Classifier* dan *Support Vector Machine*, algoritma *Xgboost Classifier* memiliki *precision* yang bernilai negative sebesar 100%, sedangkan hasil *precision* yang bernilai positive sebesar 99%, *accuracy* yang didapat sebesar 99%. Algoritma *Decesion Tree Classifier* hasil *precision negative* sebesar 50%, sedangkan untuk *precision positive* sebesar 99%, *accuracy* yang didapat sebesar 97%. Algoritma *Support Vector Machine* (SVM) hasil *precision negative* sebesar 100%, sedangkan untuk *precision positive* sebesar 98%, hasil *recall negative* 6%, *accuracy* yang didapat sebesar 98%. Sehingga dapat disimpulkan pada penelitian ini bahwa algoritma terbaik yang dapat menentukan *sentimen analysis review* hotel adalah algoritma *Xgboost* dengan *accuracy* sebesar 99%.

Diharapkan kedepannya *traveler* dapat lebih mudah dalam menentukan, memilih hotel mana yang terbaik sesuai dengan hasil *review* yang diberikan oleh pengunjung sebelumnya sehingga meminimalisir mendapatkan hotel yang memiliki *review* dan rating yang buruk.

REFERENSI

- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. <https://doi.org/10.1109/MCI.2014.2307227>
- Chen, T., He, T., & Benesty, M. (2018). XGBoost : eXtreme Gradient Boosting. R Package Version 0.71-2, 1–4.
- Eilertz, D., Mitterer, M., & Buescher, J. M. (2022). AutomRm: An R Package for Fully Automatic LC-QQQ-MS Data Preprocessing Powered by Machine Learning. *Analytical Chemistry*, 94(16), 6163–6171. <https://doi.org/10.1021/acs.analchem.1c05224>
- Farisi, A. A., Sibaroni, Y., & Faraby, S. Al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742-6596/1192/1/012024>
- Fathullah, N. S., Sari, Y. A., & Adikara, P. P. (2020). Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan Metode Klasifikasi Naïve Bayes dengan Fitur Lexicon-Based. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(2), 590–593. <http://j-ptiik.ub.ac.id>
- García-Pablos, A., Cuadros, M., & Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. *Information Technology and Tourism*, 16(1), 45–69. <https://doi.org/10.1007/s40558-015-0047-7>
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226–235. <https://doi.org/10.31311/ji.v6i2.5129>
- Peng, H. gang, Zhang, H. yu, & Wang, J. qiang. (2018). Cloud decision support model for selecting hotels on TripAdvisor.com with probabilistic linguistic information. *International Journal of Hospitality Management*, 68(October 2017), 124–138. <https://doi.org/10.1016/j.ijhm.2017.10.001>
- Saleh, A., Indra, E., & Harahap, M. (2020). Kombinasi Jaringan Learning Vector Quantization Dan Normalized Cross Correlation Pada Pengenalan Wajah. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 3(2), 13–20.

<https://doi.org/10.34012/jusikom.v3i2.851>

- Shuai, Q., Huang, Y., Jin, L., & Pang, L. (2018). Sentiment Analysis on Chinese Hotel Reviews with Doc2Vec and Classifiers. *Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018*, Iaeac, 1171–1174. <https://doi.org/10.1109/IAEAC.2018.8577581>
- Sitanggang, R., Ryan, D., Sitompul, H., Sinurat, S. H., Stumorang, A., Ziegel, D. J., Rahmad, J., & Indra, E. (2022). Sentiment Analysis Compare Linear Regression And Decision Tree Regression Algorithm To Determine Film Rating Accuracy. *10(2)*, 880–890.
- Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27(2)*, 130–135. <https://doi.org/10.11919/j.issn.1002-0829.215044>
- Taufik, A. (2018). Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel. *Jurnal Teknik Komputer*, *IV(2)*, 112–118. <https://doi.org/10.31294/jtk.v4i2.3461>
- Tummers, J., Catal, C., Tobi, H., Tekinerdogan, B., & Leusink, G. (2020). Coronaviruses and people with intellectual disability: an exploratory data analysis. *Journal of Intellectual Disability Research*, *64(7)*, 475–481. <https://doi.org/10.1111/jir.12730>