

DAFTAR ISTILAH DAN SINGKATAN

Istilah	Deskripsi	Halaman pertama kali digunakan
NLP	: <i>Natural Language Processing</i> (NLP) merupakan metode untuk melakukan pemrosesan terhadap data tekstual yang tidak terstruktur agar diubah ke bentuk yang dimengerti oleh mesin.	5
SVM	: Support Vector Machine (SVM) merupakan algoritma <i>machine learning</i> yang melakukan klasifikasi terhadap data-data yang digunakan pada penelitian ini.	5
SMOTE	: <i>Synthetic Minority Over-Sampling Technique</i> (SMOTE) merupakan teknik <i>resampling</i> yang digunakan untuk mengatasi ketidakseimbangan data dengan cara melakukan <i>oversampling</i> terhadap kelas minoritas agar seimbang dengan kelas mayoritas.	6
SMOTE-ENN	: <i>Synthetic Minority Over-Sampling Technique-Edited Nearest Neighbors</i> (SMOTE-ENN) merupakan teknik <i>resampling</i> hibrida antara <i>oversampling</i> dan <i>undersampling</i> yang digunakan untuk mengatasi ketidakseimbangan data dengan cara melakukan <i>oversampling</i> terlebih dahulu dan setelah itu melakukan <i>undersampling</i> untuk menghapus data <i>noise</i> .	6
TF-IDF	: <i>Term Frequency – Inverse Document Frequency</i> (TF-IDF) merupakan proses vektorisasi data untuk mengubah data tekstual ke dalam bentuk data numerik yang dapat diterima oleh model klasifikasi SVM.	15
TP	: <i>True Positive</i> (TP) merupakan komponen dalam <i>confusion matrix</i> yang menunjukkan jumlah hasil klasifikasi label positif pada data aktual yang memiliki label positif.	24
FP	: <i>False Positive</i> (FP) merupakan komponen dalam <i>confusion matrix</i> yang menunjukkan jumlah hasil klasifikasi label positif pada data aktual yang memiliki label negatif.	24
TN	: <i>True Negative</i> (TN) merupakan komponen dalam <i>confusion matrix</i> yang menunjukkan jumlah hasil klasifikasi label negatif pada data aktual yang memiliki label negatif.	24
FN	: <i>False Negative</i> (FN) merupakan komponen dalam <i>confusion matrix</i> yang menunjukkan jumlah hasil klasifikasi label negatif pada data aktual yang memiliki label positif.	24

Istilah	Deskripsi	Halaman pertama kali digunakan
ROC Curve	: <i>Receiver Operating Characteristic Curve (ROC Curve)</i> merupakan salah satu metode evaluasi performa model klasifikasi.	25
AUC	: <i>Area Under Curve (AUC)</i> merupakan luas keseluruhan di bawah kurva ROC.	26
RUS	: <i>Random Undersampling (RUS)</i> merupakan teknik <i>resampling</i> yang digunakan untuk mengatasi ketidakseimbangan data dengan cara melakukan <i>undersampling</i> terhadap kelas mayoritas agar seimbang dengan kelas minoritas.	30
SMOTE-Tomek	: <i>Synthetic Minority Over-Sampling Technique-Tomek Links (SMOTE-Tomek)</i> merupakan teknik <i>resampling</i> hibrida antara <i>oversampling</i> dan <i>undersampling</i> yang digunakan untuk mengatasi ketidakseimbangan data dengan cara melakukan <i>oversampling</i> terlebih dahulu dan setelah itu melakukan <i>undersampling</i> untuk menghapus data yang tumpang tindih.	32