

Perbandingan kinerja Support Vector Machine (SVM), Decision Tree, dan k-Nearest Neighbors untuk memprediksi adiksi internet dan kondisi kesehatan mental: studi kasus mahasiswa Saintek Universitas Indonesia = Performance comparison of Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors for internet addiction and mental health status prediction: a study case of STEM Students at Universitas Indonesia

Valentinus Paramarta, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20502761&lokasi=lokal>

Abstrak

Semakin tinggi penetrasi penggunaan Internet seseorang, maka akan semakin berpotensi terkena Gangguan Adiksi Internet (GAI) yang dapat berdampak buruk pada status kesehatan mental penggunanya. Mayoritas penduduk Indonesia telah menggunakan layanan Internet selama 2 sampai 3 tahun dengan penggunaan rata-rata di atas 8 jam perhari. Hal tersebut menunjukkan penggunaan Internet dan potensi dampaknya pada kesehatan mental di Indonesia penting untuk diperhatikan sedini mungkin. Penelitian lain menunjukkan bahwa tingkat kesehatan mental yang dialami seseorang dapat mempengaruhi perilaku penggunaan Internetnya, sehingga menyebabkan munculnya keinginan yang tidak terkendali dan berlebihan dalam mengakses Internet. Secara tidak langsung, hal tersebut menyatakan bahwa kesehatan mental seseorang juga dapat diamati melalui tingkah laku serta kebiasaan seseorang dalam menggunakan Internet. Prediksi GAI dan gangguan kesehatan mental mahasiswa UI dilakukan dengan menggunakan algoritma pembelajaran mesin Support Vector Machine (SVM) berdasarkan perilaku penggunaan Internet yang dilakukan. Sampel diambil dari mahasiswa UI rumpun Ilmu Saintek (Ilmu Komputer, Teknik, dan MIPA). Data yang diambil adalah riwayat penelusuran halaman website yang diakses oleh mahasiswa dan hasil kuesioner Internet addiction test (IAT) dan General Health Questionnaire (GHQ-12). Riwayat penelusuran website dijadikan himpunan fitur yang merepresentasikan perilaku penggunaan Internet responden, sedangkan hasil skor kuesioner IAT dan GHQ-12 digunakan untuk menjadi ground truth atau label pada dataset. Tahapan preprocessing yang dilakukan adalah metode Synthetic Minority Over-Sampling Technique (SMOTE) untuk mengatasi ketidakseimbangan persebaran data pada kelas data yang digunakan. Metode SVM selanjutnya dibandingkan dengan performa lainnya seperti Decision Tree dan k-Nearest Neighbor (kNN). Untuk meningkatkan performa akurasi, peneliti menggunakan metode grid search untuk mendapatkan parameter terbaik. Proses validasi dilakukan menggunakan cross-validation pada metode grid search. Hasil yang didapatkan menunjukkan bahwa performa akurasi tertinggi pada SVM untuk memprediksi GAI adalah 88% pada dataset kedua. Saat dilakukan perbandingan hasil dengan metode pembelajaran mesin Decision Tree dan kNN, didapatkan performa nilai akurasi tertinggi dicapai pada metode Decision Tree dengan nilai akurasi sebesar 96%. Sedangkan untuk prediksi gangguan kesehatan mental, metode SVM mendapatkan nilai performa akurasi tertinggi sebesar 71% pada dataset gabungan. Saat dilakukan perbandingan hasil performa akurasi dengan Decision Tree dan kNN, didapatkan nilai performa akurasi tertinggi dicapai pada metode kNN sebesar 72%. Hasil penelitian ini menunjukkan bahwa metode grid search meningkatkan performa SVM, Decision Tree, dan

kNN karena adanya perubahan nilai parameter.

.....Excessive internet usage lead to potential Internet Addiction Disorders (IAD) which affect user`s mental health. The majority of Indonesian people have been used Internet services for 2 until 3 years in their lives with an average use of above 8 hours per day. It shows that an increase of internet usage has a positive potential impact to an increase in mental disorder. Other research shows that the level of mental health experienced by a person can influence his Internet usage behavior, thus causing an uncontrolled and excessive desire to access the Internet. It could be concluded that the mental health can also be observed through one`s behavior and habits in using the Internet. This study predicts the internet addiction disorder (IAD) and mental health disorder status of UI students by using machine learning based on Support vector Machine (SVM) algorithm. This study used behaviour of internet usage for the input. Samples used in this study were taken from Universitas Indonesia`s students with Science and Technology background. The data collection period was set before and after the exam period. Data collected in this study included history of website accessed by students and questionnaires based on Internet addiction test (IAT) and General Health Questionnaire (GHQ-12). Student`s website history would be used as feature data set that represent user internet usage behavior, while the IAT and GHQ-12 questionnaires results were used as the label. The preprocessing stage was carried out using Synthetic Minority Over-Sampling Technique (SMOTE) method to overcome the imbalance of data distribution in class used. Then, student`s website history would be analyzed using machine learning based on SVM algorithm to predict IAT and mental health status. This study also compared other algorithms such as Decision Tree and k-Nearest Neighbor (kNN). The optimization of machine learning model was conducted using grid search method to obtain the best parameters. The validation of the model would be carried out using the cross-validation obtained from grid search method. Based on the results obtained, it shows that the highest accuracy for predicting internet addiction was obtained from SVM algorithm with 88% accuracy for the second dataset. Comparison with other models showed that Decision Tree obtained the highest accuracy value of 96% for predicting internet addiction. For the prediction of mental health disorder, SVM algorithm obtained the highest accuracy than Decision Tree or kNN. The SVM algorithm can predict with accuracy of 71% with combined dataset. When comparing the accuracy result with the accuracy of Decision Tree and kNN, the highest accuracy value of 72% was achieved by kNN method. The optimal value of accuracy is obtained when the grid search method is performed. The results of this study indicate that the grid search method has succeeded in improving the performance of SVM, Decision Tree, and kNN due to parameter value changes.