

Pendekatan machine learning untuk klasifikasi fitur yang memengaruhi popularitas sebuah makalah penelitian = Machine learning approach in classifying features determining the popularity of a published paper

Alexander Patrick, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20528055&lokasi=lokal>

Abstrak

Popularitas sebuah paper yang terpublikasi dapat dilihat dari jumlah sitasi yang diperoleh paper tersebut. Akan tetapi, faktor-faktor yang mendukung banyak atau sedikitnya jumlah sitasi yang diperoleh bisa bermacam-macam. Faktor-faktor tersebut merupakan fokus pencarian pada penelitian ini. Pendekatan machine learning digunakan untuk mengetahui faktor-faktor tersebut. Beberapa fitur telah diekstrak dari dataset yang berisi kumpulan paper. Metode klasifikasi digunakan dalam supervised learning dengan model yang dibentuk dari dataset yang digunakan. Algoritma Logistic Regression dipakai untuk melakukan fitting terhadap model dengan hasil daya diskriminasi sistem sebesar 74,1% yang dilihat dari luas wilayah di bawah kurva ROC (Area Under Curve/AUC). Nilai koefisien dari model Logistic Regression digunakan sebagai feature importance untuk mencari nilai pengaruh dari tiap fitur terhadap output klasifikasi baik positif maupun negatif.

.....Popularity of a published paper can be indicated by its citation number. However, the factors determining the number of citation may vary. Those factors are the focus of this research. A machine learning approach is used to find out the factors. Some features are going to be extracted from a dataset of published papers. A classification method is going to be applied in a supervised learning with the machine learning model extracted from the dataset. A classification algorithm Logistic Regression is used to fit the model resulting a discrimination power of 74.1% from a calculation of area under ROC curve (AUC). A feature importance approach using coefficient score from Logistic Regression is also applied in determining the importance of each feature in determining the negative and positive classification.