

## ABSTRAK

Kanker paru sebagai satu dari tiga penyakit yang paling mematikan di dunia serta memiliki perkembangan yang sangat pesat. Berdasarkan hal ini, peneliti melakukan penelitian untuk dapat memprediksi faktor yang mempengaruhi kanker paru. Salah satu metode untuk mengidentifikasi hal ini adalah dengan menggunakan metode *data mining* dengan menggunakan teknik klasifikasi. Peneliti menggunakan beberapa algoritma populer dalam klasifikasi untuk melakukan perbandingan terhadap algoritma yang paling akurat untuk klasifikasi kanker paru. Algoritma yang digunakan diantaranya *K-Nearest Neighbour*, *Random Forest Classifier*, *Logistic Regression*, *Linear SVM*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, *Kernel SVM* dan *MLPClassifier*. Peneliti menggunakan algoritma tersebut karena pada penelitian yang peneliti temukan di *platform kaggle* meneliti tentang perbandingan algoritma tersebut menggunakan *dataset breast cancer*. Pada penelitian sebelumnya, ada peneliti yang menggunakan SVM dan mendapat nilai akurasi sebesar 96,47%, Neural Network sebesar 97,06%, dan memakai Naïve Bayes dengan akurasi 91,18% untuk meneliti kanker payudara. Adapun yang menjadi perbedaan dari penelitian sebelumnya yaitu pada penelitian ini, menggunakan beberapa algoritma yang ada pada *Machine Learning* seperti *K-Nearest Neighbour*, *Random Forest Classifier*, *Logistic Regression*, *Linear SVM*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, *Kernel SVM* dan *MLPClassifier*. Penelitian ini dilakukan untuk melihat apakah hasil akurasi algoritma yang peneliti lakukan dengan menggunakan *dataset lung cancer* memiliki hasil yang berbeda atau tidak. Hasil dari penelitian ini menemukan bahwa algoritma yang lebih akurat adalah *Random Forest* dan *Gradient Boosting* dengan nilai akurasi sebesar 100% sedangkan pada penelitian sebelumnya juga sama namun *Gradient Boosting* lebih tinggi nilai akurasinya dibanding *Random Forest*. Kemudian, berdasarkan data yang digunakan pada penelitian ini, faktor yang paling mempengaruhi dalam memprediksi didiagnosa terkena kanker paru-paru adalah obesitas dan batuk berdarah.