

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Paru-paru adalah salah satu organ dari sistem pernapasan yang berfungsi sebagai tempat terjadinya pertukaran karbondioksia dan oksigen dalam darah. Permasalahan yang sering terjadi pada sistem pernapasan adalah kualitas udara yang telah tercemar, sehingga udara yang dihirup mengandung banyak bakteri yang dapat menyerang sistem pernapasan terutama paru-paru [1]. Salah satu penyakit yang menyerang sistem pernapasan yaitu kanker paru-paru.

Kanker paru-paru merupakan penyakit dengan ciri khas adanya pertumbuhan sel yang tidak terkontrol pada jaringan paru-paru. Bila tidak dirawat pertumbuhan sel ini dapat menyebar keluar dari paru-paru melalui suatu proses yang disebut *metastasis* ke jaringan yang terdekat atau bagian tubuh yang lainnya. Kanker paru sebagai satu dari tiga penyakit yang paling mematikan di dunia memiliki perkembangan yang sangat pesat. Beberapa faktor yang mempengaruhi pesatnya perkembangan penyakit kanker paru di dunia antara lain adalah paparan dalam jangka waktu yang lama terhadap asap tembakau, faktor genetik, gas radon dan polusi udara.[2]

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*. *Data mining* adalah salah satu cabang ilmu di bidang komputer, dimana perkembangannya sangat pesat. *Data mining* telah banyak diaplikasikan ke dalam berbagai bidang seperti ilmu pengetahuan, teknik dan bisnis. Tujuan utama dari *data mining* adalah untuk menemukan informasi penting yang terdapat di dalam *database* yang sangat besar. Beberapa teknik data mining telah banyak dikembangkan seperti teknik asosiasi, *clustering*, *sequence pattern* dan klasifikasi. Dalam hal ini, peneliti menggunakan teknik klasifikasi dikarenakan *dataset* yang peneliti temukan merupakan data tentang klasifikasi dan sesuai dengan tujuan penelitian ini.

Namun, dalam melakukan klasifikasi faktor yang mempengaruhi kanker paru-paru menggunakan beberapa algoritma yang paling populer dalam klasifikasi untuk melakukan perbandingan terhadap algoritma mana yang lebih akurat untuk klasifikasi kanker paru ini. Adapun algoritma yang digunakan diantaranya Algoritma *K-Nearest Neighbour*, *Random Forest Classifier*, *Logistic Regression*, *Linear SVM*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, *Kernel SVM* dan *MPLClassifier*. Alasan menggunakan beberapa algoritma tersebut, dikarenakan peneliti menemukan sebuah penelitian di *platform Kaggle* tentang perbandingan algoritma tersebut menggunakan *dataset breast cancer*. Oleh karena itu, peneliti tertarik untuk melakukan penelitian untuk mengetahui apakah hasil akurasi algoritma dengan menggunakan *dataset lung cancer* sama dengan hasil akurasi algoritma yang ada pada penelitian tersebut atau berbeda. Berdasarkan uraian-uraian diatas, penulis mengangkat judul “**Optimalisasi Metode Klasifikasi Kanker Paru Menggunakan Machine Learning Berbasis EDA**”.

## 1.2 Rumusan Masalah

Rumusan masalah berdasarkan latar belakang yang telah diuraikan adalah sebagai berikut :

1. Algoritma manakah yang lebih optimal/akurat dalam klasifikasi kanker paru-paru?
2. Parameter manakah yang paling mempengaruhi dalam memprediksi seseorang di diagnosa terkena kanker paru-paru?

## 1.3 Batasan Masalah

Batasan masalah yang terdapat dalam penelitian ini antara lain :

1. Penelitian ini menggunakan *dataset lung cancer*.
2. Pada penelitian ini, *dataset* yang digunakan diambil dari *Kaggle* yang merupakan *platform* penyedia data dengan jumlah yang banyak, yang mana semakin banyak jumlah data yang ada, maka akan semakin baik dalam melakukan penelitian.

3. Algoritma yang digunakan dalam penelitian ini antara lain Algoritma *K-Nearest Neighbour*, *Random Forest Classifier*, *Logistic Regression*, *Linear SVM*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, *Kernel SVM* dan *MLPClassifier*.
4. Peneliti menggunakan *Machine Learning* dengan bahasa pemrograman *Phyton*.

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah :

1. Untuk mengetahui dari beberapa algoritma yang digunakan, algoritma yang mana yang paling akurat dalam klasifikasi kanker paru.
2. Untuk mengetahui parameter apa yang paling mempengaruhi dalam memprediksi seseorang di diagnosa terkena kanker paru-paru.

#### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah :

1. Untuk mendapatkan pola *sequence* yang berguna dalam data rekam medis pasien untuk membantu diperolehnya klasifikasi tentang kanker paru-paru.
2. Mendapatkan informasi gambaran dari parameter mana yang paling mempengaruhi dalam memprediksi seseorang di diagnosa terkena kanker paru-paru.
3. Bagi penulis, untuk menambah pengetahuan terutama dalam menganalisis suatu permasalahan menggunakan algoritma yang ada didalam *data mining* dengan menggunakan Algoritma *K-Nearest Neighbour*, *Random Forest Classifier*, *Logistic Regression*, *Linear SVM*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, *Kernel SVM* dan *MLPClassifier*.

#### **1.6 Sistematika Penulisan**

Sistematika penulisan penelitian ini dibagi menjadi 4 bab, yaitu :

## **BAB 1 PENDAHULUAN**

Bab ini berisi penjelasan mengenai judul penelitian, latar belakang pemilihan judul, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

## **BAB 2 METODE PENELITIAN**

Bab ini membahas tentang metode yang digunakan dalam penelitian ini dan apa saja langkah-langkah yang dilakukan dengan menggunakan metode tersebut.

## **BAB 3 HASIL DAN PEMBAHASAN**

Bab ini membahas hasil dari penelitian dan pengujian terhadap algoritma yang digunakan yang digunakan yaitu Algoritma *K-Nearest Neighbour*, *Random Forest Classifier*, *Logistic Regression*, *Linear SVM*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boosting*, *Kernel SVM* dan *MPLClassifier*.

## **BAB 4 KESIMPULAN DAN SARAN**

Bab ini menjelaskan hasil akhir dari semua penulisan yang dikerjakan yaitu kesimpulan dan saran yang berisi masukan-masukan untuk penelitian-penelitian yang akan dilakukan selanjutnya.