

BAB I

PENDAHULUAN

1.1. Latar Belakang

Hepatocellular Carcinoma (HCC) atau biasa disebut kanker hati merupakan salah satu tumor ganas dan heterogen tanpa pengobatan yang efektif serta merupakan jenis kanker yang paling fatal diantara jenis kanker lainnya. Faktor resiko pada HCC adalah infeksi kronik Hepatitis B, Hepatitis C dan sirosis hati serta faktor keturunan. Risiko seseorang yang terkena HCC dengan infeksi Hepatitis B kronik adalah 100 kali lebih banyak daripada orang yang tidak terinfeksi. Di wilayah dengan angka penderita HCC tinggi, rasio kasus laki-laki dan perempuan dapat mencapai 8:1. Hal tersebut dikarenakan laki-laki lebih banyak yang terpapar faktor-faktor resiko HCC, seperti virus Hepatitis dan alkohol (Sulaiman, 2012).

Hepatitis B merupakan suatu sindroma klinis atau patologis yang ditandai dengan berbagai tingkat peradangan dan nekrosis pada hepar yang disebabkan oleh virus Hepatitis B yang menyerang sel hati, dimana infeksi tersebut dapat berlangsung secara akut atau kronik. Mekanisme terjadinya Hepatitis B akut atau kronik dan HCC diawali oleh rusaknya sel hati. Berdasarkan berbagai penelitian yang telah dilakukan mengungkapkan bahwa faktor penderita (umur, jenis kelamin, faktor genetik dan *imunologik*) serta respon imun seluler terhadap virus Hepatitis B terlibat dalam klirens virus dan bertanggung jawab atas terjadinya HCC.

Singapura merupakan negara dengan tingkat produksi minuman beralkohol tertinggi di Asia dan merupakan surga bagi industri minuman beralkohol di Asia.

Singapura memanfaatkan ekspatriat dan wisatawan guna mengembangkan bisnis tersebut. Seperti yang diketahui bahwa Singapura memiliki budaya yang lebih condong ke Barat, maka negara tersebut memiliki tingkat konsumsi minuman beralkohol yang tinggi. Minuman beralkohol tersebut sangat berpengaruh terhadap kesehatan, salah satunya yaitu menjadi penyebab munculnya penyakit kanker hati atau HCC.

Penyakit HCC di Amerika Serikat menjadi penyebab kematian akibat kanker terbanyak ke-9. HCC banyak dijumpai pada penderita sirosis, Hepatitis B kronis, NASH, dan obesitas. Diperkirakan bahwa ada sekitar 841.000 kasus baru dan 782.000 kematian yang disebabkan oleh kanker hati di seluruh dunia pada tahun 2018 (Bray, 2018), dengan jumlah pasien di China lebih dari setengah pasien HCC global (Jemal et al., 2011; Bray, 2018). Distribusi global dari HCC erat kaitannya dengan prevalensi geografis VHB dan VHC kronik yang jumlahnya mencapai 400 juta penduduk di seluruh dunia. Secara epidemiologi, jumlah kasus VHB kronik adalah sebesar 52% kasus. Penderita VHB kronik memiliki resiko 102 kali lebih tinggi daripada yang bukan kronik untuk terjadinya HCC.

Meskipun banyak penelitian yang telah membahas tentang penyakit HCC terinfeksi Hepatitis B, akan tetapi masih banyak yang belum diketahui tentang penyakit tersebut. Hal utama yaitu tentang perubahan biologis yang menyebabkan terjadinya HCC terinfeksi Hepatitis B, penyebab penyakit ini berlangsung lebih cepat pada beberapa orang dan cara mencegah atau menghentikan penyakit tersebut. Beberapa kasus dari penyakit tersebut terdapat pasien dengan jaringan tumor maupun normal berdekatan. Seperti yang telah dijelaskan bahwa HCC

merupakan jenis tumor ganas, akan tetapi pada sebagian kasus terdapat pasien yang memiliki jaringan normal berdekatan. Jaringan tersebut mirip dengan jaringan hati yang normal, akan tetapi terdapat pola tumor padat. Pada kedua jaringan tersebut tentunya memiliki status gen yang berbeda, maka untuk membandingkan atau melihat kelas dari dua perbedaan jaringan ini perlu adanya suatu analisis untuk mendiagnosis penyakit tersebut. Banyak peneliti yang mempercayai bahwa kunci untuk menangani penyakit ini adalah dengan deteksi dini. Deteksi tersebut dapat diketahui dengan melihat ekspresi dari gen yang terkandung dalam DNA. Ilmu yang mempelajari tentang biologis dengan melibatkan suatu sistem yaitu bioinformatika.

Fatchiyah (2009) menyatakan bahwa bioinformatika dapat dikatakan sebagai ilmu yang mempelajari tentang penerapan teknik komputasi untuk menganalisis berbagai informasi biologi. Bioinformatika telah banyak diterapkan pada berbagai bidang, contohnya yaitu untuk pengelolaan informasi data biologi, melakukan prediksi struktur bentuk protein maupun *Ribonucleic Acid* (RNA), analisis data ekspresi gen, perakitan ratusan pasang basa DNA menjadi satu sekuen genom utuh dengan ukuran mencapai milyaran pasang basa (*genome assembling*), melakukan prediksi keberadaan gen dan perangkat pelengkapannya seperti promotor dalam genom hingga memprediksi fungsinya.

Salah satu teknologi yang berkembang dalam bidang bioinformatika adalah teknologi *microarray*, yang merupakan sebuah perangkat berupa chip yang didalamnya berisi ribuan gen. Pada dasarnya, analisis *microarray* merupakan proses analisis pada ekspresi gen (*gene expression*) melalui *microarray*. Ekspresi

gen merupakan proses untuk menterjemahkan informasi dari gen menjadi produk gen yang berguna, yaitu protein (Sanchez & M. Carme Ruiz de Villa, 2008). Pada data *microarray*, dapat digunakan untuk mendeteksi atau mengklasifikasikan jaringan suatu penyakit pada manusia. Data *microarray* menghasilkan ekspresi gen yang berisi informasi gen-gen yang ada dan kemudian dicocokkan dengan suatu penyakit (Trevino, Falciani, & Barrera-Saldana, 2007).

Analisis yang dapat digunakan untuk melihat adanya perbedaan dari dua kelas pada penyakit tersebut adalah analisis klasifikasi. Klasifikasi merupakan proses menyatakan suatu objek data sebagai salah satu dari sebuah kategori (kelas) yang telah didefinisikan sebelumnya (Zaku., et al. 2013). Klasifikasi dapat digunakan untuk berbagai aplikasi, diantaranya adalah sebagai deteksi kecurangan, pengelolaan pelanggan, diagnosis pada bidang medis, prediksi pada penjualan dan lainnya. Analisis klasifikasi tersebut dapat menghasilkan sebuah model, akurasi dan juga *lost (error)*.

Metode klasifikasi dapat dibangun menggunakan teknik pembelajaran pada bidang *Machine Learning*. Teknik komputasi *machine learning* dapat digunakan untuk menganalisis pemilihan gen ataupun protein yang memiliki sifat terkait dan mengklasifikasikan tipe sampel ekspresi gen pada data *microarray* (Yang, Yang, Zhou, & Zomaya, 2016). *Machine Learning* adalah suatu sistem yang dibangun berdasarkan pembelajaran pada data-data besar. Algoritma yang digunakan pada *machine learning* adalah *supervised learning*, dimana algoritma tersebut digunakan untuk pengamatan dengan hasil yang diperoleh telah diketahui kelasnya. Algoritma pemodelan pada *supervised learning* yaitu klasifikasi, dimana pada penelitian ini

menggunakan metode *Multilayer Perceptron* (MLP) dan *Xtreme Gradient Boosting* (XGBoost).

Multilayer Perceptron (MLP) merupakan suatu algoritma pada pembuatan model *neural network* atau jaringan saraf tiruan. Jaringan syaraf tiruan memiliki kemampuan yang baik dalam menganalisa pola data. Kemampuan ini menjadi salah satu alasan mengapa jaringan syaraf tiruan banyak dipilih sebagai metode dalam melakukan prediksi. MLP merupakan *perceptron* dengan satu atau lebih *hidden layer*, dimana input sebuah perceptron adalah output dari perceptron sebelumnya. Tidak seperti *perceptron* yang hanya dapat memodelkan permasalahan linear, MLP juga dapat menyelesaikan permasalahan non-linear (Suyanto, Data Mining , 2017). MLP adalah *classifier linier* yang menggunakan teknik *supervised learning*, untuk mengubah himpunan *input* menjadi himpunan *output*. Klasifikasi ini didasarkan pada fungsi prediktor linier, yang menggabungkan satu set bobot dengan vektor masukan. Dasar dari jaringan ini adalah ketika data ditampilkan pada lapisan *input*, *node-node* jaringan akan melakukan perhitungan pada lapisan dengan berturut-turut sampai nilai *output* diperoleh. Sinyal *output* ini akan menunjukkan *input* data dengan tepat dan diharapkan nilai *output* memiliki nilai yang tinggi (Yan, et al. 2006).

Xtreme Gradient Boosting (XGBoost) merupakan suatu metode pada *machine learning*, dimana XGBoost merupakan algoritma regresi dan klasifikasi dengan metode *ensemble* yang merupakan suatu varian dari algoritma *Tree Gradient Boosting* yang dikembangkan dengan optimasi 10 kali lebih cepat dibandingkan *Gradient Boosting* lainnya (Chen & Guestrin, 2016). Pembangunan

model dilakukan dengan menggunakan metode *boosting*, yaitu dengan membuat model baru untuk memprediksi *error/residual* dari model sebelumnya. Model baru ditambahkan hingga tidak ada lagi perbaikan pada *error* yang dapat dilakukan.

Penelitian tentang *Multilayer Perceptron* (MLP) telah dilakukan beberapa kali, seperti penelitian yang dilakukan oleh (Naf'an & Arifin, 2017) yang berjudul “Identifikasi Tanda Tangan Berdasarkan *Grid Entropy* Menggunakan *Multi Layer Perceptron*” dengan hasil yang diperoleh menunjukkan bahwa pengujian terbaik adalah untuk pengujian ukuran *grid* 8x8 dan menggunakan *citra outline*, yaitu dengan tingkat akurasi sebesar 97,78%, nilai korelasi 0,981 dan nilai kappa 0,977. Selanjutnya penelitian yang dilakukan oleh (Ardilla, Tjandrasa, & Arieshanti, 2014) yang berjudul “Deteksi Penyakit *Epilepsi* dengan Menggunakan Entropi Permutasi, *K-means Clustering* dan *Multilayer Perceptron*” dengan hasil yang diperoleh menunjukkan bahwa model klasifikasi *multilayer perceptron* akan menghasilkan performa maksimal ketika dilakukan eksekusi pada algoritma entropi permutasi dengan jumlah dimensi antara 3 hingga 6.

Penelitian tentang *Xtreme Gradient Boosting* (XGBoost) juga telah beberapa kali dilakukan, seperti penelitian yang dilakukan oleh (Prasetyo, Christianto, & Hartomo, 2019) berjudul “Analisis Data Citra Landsat 8 OLI Sebagai Indeks Prediksi Kekeringan Menggunakan *Machine Learning* di Wilayah Kabupaten Boyolali dan Purworejo” dengan hasil yang diperoleh menunjukkan bahwa metode XGBoost lebih baik digunakan dibandingkan dengan metode Random Forest karena mempunyai nilai akurasi dan nilai kappa yang lebih tinggi. Selanjutnya adalah penelitian yang dilakukan oleh (Maalik, Kusuma, & Wahjuni, 2019)

berjudul “*Comparison Analysis of Ensemble Technique With Boosting (XGBOOST) and Bagging (Random Forest) for Classify Splice Junction DNA Sequence Category*” dengan hasil yang diperoleh menunjukkan bahwa nilai akurasi sebesar 96,24% untuk metode XGBoost dan 95,11% untuk metode Random Forest.

Setelah mempelajari dan memahami beberapa penelitian terdahulu yang berkaitan dengan metode dan objek yang digunakan pada penelitian ini, maka dapat diketahui perbedaan yang dimiliki dari penelitian ini dengan penelitian-penelitian sebelumnya yang terletak pada objek yang digunakan dengan membandingkan dua metode klasifikasi. Objek yang digunakan pada penelitian ini adalah data ekspresi gen HCC yang memiliki infeksi virus Hepatitis B kronis yang diklasifikasikan menggunakan 2 metode klasifikasi yaitu *Multilayer Perceptron* (MLP) dan *Xtreme Gradient Boosting* (XGBoost) yang kemudian hasil dari kedua metode tersebut dibandingkan metode mana yang paling baik dalam memprediksi gen jaringan tumor dan jaringan normal berdekatan pada HCC yang terinfeksi virus Hepatitis B kronik.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas maka penulis dapat memaparkan beberapa rumusan masalah yang akan diselesaikan pada penelitian, yaitu:

1. Bagaimana hasil klasifikasi jaringan tumor dan jaringan normal berdekatan pada pasien HCC terinfeksi Hepatitis B kronik dengan menggunakan metode *Multilayer Perceptron*?

2. Bagaimana hasil klasifikasi jaringan tumor dan jaringan normal berdekatan pada pasien HCC terinfeksi Hepatitis B kronik dengan menggunakan metode *Xtreme Gradient Boosting*?
3. Manakah metode yang terbaik dalam mengklasifikasikan jaringan tumor dan jaringan normal berdekatan pada pasien HCC terinfeksi Hepatitis B kronik?

1.3. Tujuan Penelitian

Berdasarkan rumusan yang telah dipaparkan, maka tujuan dari penelitian adalah sebagai berikut:

1. Mengklasifikasikan jaringan tumor dan jaringan normal berdekatan pada pasien HCC terinfeksi Hepatitis B kronik dengan menggunakan metode *Multilayer Perceptron*.
2. Mengklasifikasikan jaringan tumor dan jaringan normal berdekatan pada pasien HCC terinfeksi Hepatitis B kronik dengan menggunakan metode *Xtreme Gradient Boosting*.
3. Memperoleh metode yang terbaik dalam mengklasifikasikan jaringan tumor dan jaringan normal berdekatan pada pasien HCC terinfeksi Hepatitis B kronik.

1.4. Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Manfaat Teoritis

Membantu perkembangan ilmu pengetahuan mengenai metode klasifikasi sehingga dapat digunakan dapat digunakan sebagai bahan bacaan dan referensi

bagi pembaca dalam melakukan analisis terutama pada metode *Multilayer Perceptron* dan *Xtreme Gradient Boosting* pada ekspresi gen.

2. Manfaat Praktis

a. Bagi Peneliti

Peneliti mampu menerapkan metode yang sesuai dalam materi yang telah dipelajari serta peneliti memiliki pengetahuan dan wawasan mengenai klasifikasi data ekspresi gen HCC terinfeksi hepatitis B pada jaringan tumor dan normal berdekatan dengan metode *Multilayer Perceptron* dan *Xtreme Gradient Boosting*.

- ### b. Bagi pihak terkait yang diantaranya bagi rumah sakit, masyarakat, tenaga kesehatan yaitu dapat membantu mengetahui gen-gen yang paling berpengaruh pada penyakit HCC terinfeksi hepatitis B serta memudahkan tenaga kesehatan untuk mengatasi dan mendeteksi dini penyakit tersebut.

1.5. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah:

1. Melakukan klasifikasi dengan menggunakan metode *Multilayer Perceptron* dan *Xtreme Gradient Boosting*.
2. Metode *Multilayer Perceptron* menggunakan algoritma pembelajaran *Backpropagation* dengan fungsi aktivasi *Rectified Linier Unit* (ReLU).
3. Metode *Xtreme Gradient Boosting* menggunakan *Tuning Parameter Eta*, *Gamma*, *Max_depth*, *Min_child_weight*, *Subsample* dan *Colsample_bytree*.