

## **BAB III**

### **LANDASAN TEORI**

#### **3.1 *Cryptocurrency***

*Cryptocurrency* adalah metode untuk membentuk "koin" virtual dan menyediakan kepemilikan dan transaksi yang aman menggunakan masalah kriptografi. Masalah ini dirancang agar mudah diverifikasi tetapi secara komputasi sulit untuk mencapai solusi. Berbagai *cryptocurrency* menggunakan fungsi yang berbeda untuk tujuan ini, yang paling umum menjadi target hash, dimana hash dihitung sehingga datang lebih rendah dari nilai tertentu. Target hash (mis., Kesulitan masalah) disesuaikan setiap kali berdasarkan pada daya komputasi total pada jaringan, yang memiliki keuntungan menjaga waktu antara solusi lebih atau kurang konstan. Bukti kerja intensif komputasi adalah metode dimana transaksi diverifikasi sebagai unik dan dapat dipercaya. Untuk mendorong partisipasi, transaktor dapat menyertakan biaya transaksi yang masuk ke pengguna pertama yang berhasil memverifikasinya. Selain itu, jaringan menghadiahkan verifier dengan sejumlah koin setelah mereka berhasil memverifikasi satu blok transaksi. Proses ini, yang disebut penambangan. Penambangan adalah cara di mana pasokan koin pada jaringan diperluas, dan kesulitan yang dapat disesuaikan memastikan bahwa kemajuan komputasi tidak akan mempengaruhi tingkat ekspansi (Harwick, 2016).

Sistem *Cryptocurrency* umumnya mengklaim menyediakan pemrosesan transaksi anonim dan terdesentralisasi. Anonimitas ini dapat digunakan sebagai tindakan pencegahan tambahan untuk kerahasiaan dan privasi pengguna.

Penerimaan dan permintaan *cryptocurrency* telah meningkat seratus kali lipat selama beberapa tahun terakhir. Demikian pula, industri sekitar *cryptocurrency* telah berevolusi sejak awal dan sejumlah pemangku kepentingan sekarang terkait dengan perdagangan yang berkembang dan penerimaan mata uang kripto. Saat ini, *cryptocurrency* sudah tersedia di ratusan bursa di seluruh dunia terhadap mata uang fiat (Hameed & Farooq, 2016).

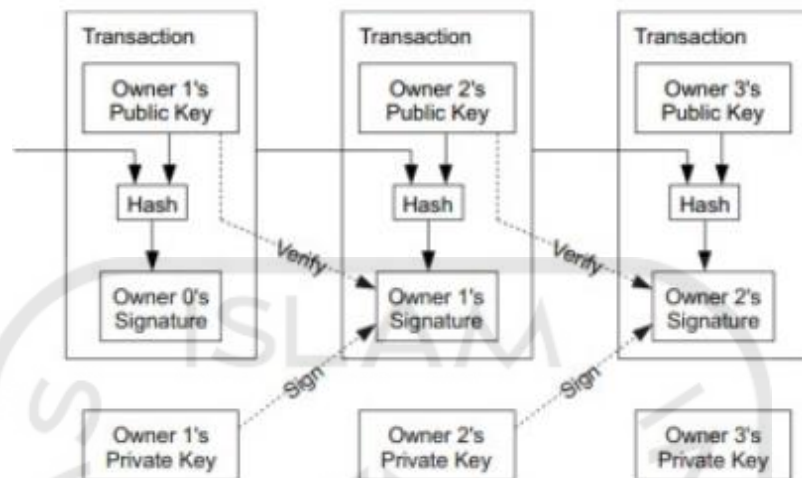
### 3.2 Bitcoin

Bitcoin atau disingkat dengan “BTC” adalah mata uang digital, yang tidak dikeluarkan oleh lembaga, organisasi ataupun pemerintah dalam regulasinya. Bitcoin memanfaatkan jaringan *peer-to-peer network* sebagai media distribusinya menggunakan protokol kriptografi canggih. Pertama diusulkan oleh Satoshi Nakamoto pada tahun 2008, *software* bitcoin dibuat dan mulai dijalankan di tahun 2009 (Nakamoto, 2008).

Sebuah nilai bitcoin (*electronic coin*) merupakan rangkaian dari sebuah tanda tangan digital. Semua transaksi Bitcoin diposting dalam blok ke buku besar terbuka yang dikenal sebagai Blockchain untuk diverifikasi oleh penambang menggunakan bukti kriptografi. Verifikasi ini terjadi dalam sistem tanpa kepercayaan dan tanpa perantara yang diperlukan untuk meneruskan dana dari pengirim ke penerima. Bitcoin menawarkan peluang baru untuk prediksi karena usianya yang relatif muda dan volatilitas yang dihasilkan. Selain itu, ia unik dalam kaitannya dengan mata uang fiat tradisional dalam hal sifatnya yang terbuka. Tidak seperti mata uang dan aset tradisional, bitcoin mudah dibawa, dapat dibagi, dan tidak dapat diubah. Bitcoin meningkatkan efisiensi sistem dan memungkinkan penyediaan layanan keuangan dengan biaya yang jauh lebih rendah, memberi pengguna lebih banyak kekuatan dan kebebasan (Blockchain, 2019).

Sebagai perbandingan, tidak ada data lengkap mengenai transaksi tunai atau uang dalam sirkulasi mata uang fiat. Hipotesis pasar efisien yang terkenal (Fama, 2010) menunjukkan harga aset seperti mata uang mencerminkan semua informasi yang tersedia, dan sebagai hasilnya berdagang dengan nilai wajarnya. Meskipun ada banyak data yang tersedia terkait dengan Bitcoin dan jaringannya, tidak semua pelaku pasar akan memanfaatkan semua informasi ini secara efektif dan akibatnya mungkin tidak tercermin dalam harga (McNally, 2016).

Pembahasan detail mengenai cara kerja bitcoin diluar dari cakupan tulisan ini. Pada Gambar 3.1 Perbandingan dapat dilihat sebuah blok rantai transaksi yang digunakan untuk menyimpan informasi berupa informasi pemilik bitcoin tersebut.



**Gambar 3.1** Blok transaksi bitcoin (Mulyanto & Mulia, 2014)

Jumlah peredaran bitcoin dapat diprediksi dengan nilai yang dibatasi sebesar 21 juta bitcoin, dan nilai 21 juta bitcoin tersebut akan dicapai sekitar tahun 2040, hingga saat tulisan ini dibuat baru sekitar 11 juta bitcoin yang telah dihasilkan dan beredar (Mulyanto & Mulia, 2014).

### 3.3 Ekonomi Bitcoin

Model Barro (1979) dan Ciaian et al. (2016) dalam jurnal yang ditulis oleh Jang dan Lee (2017) memberikan model harga Bitcoin sederhana dalam kondisi pasar yang sempurna. Dalam model ini, Bitcoin diasumsikan memiliki nilai mata uang dan dapat ditukar dengan mata uang tradisional, yang berada di bawah kendali bank sentral dan dapat digunakan untuk membeli barang dan jasa. Total pasokan Bitcoin,  $S_B$ , diwakili oleh:

$$S_B = P_B B \quad (3.1)$$

dimana  $P_B$  menunjukkan nilai tukar antara Bitcoin dan dolar (yaitu dolar per unit Bitcoin), dan  $B$  adalah total kapasitas Bitcoin yang beredar.

Total permintaan Bitcoin tergantung pada tingkat harga barang atau jasa umum,  $P$ ; ukuran ekonomi Bitcoin,  $E$ ; dan kecepatan Bitcoin,  $V$ , yang merupakan frekuensi di mana unit Bitcoin digunakan untuk membeli barang atau jasa. Total permintaan Bitcoin,  $D_B$ , digambarkan sebagai berikut:

$$D_B = \frac{PE}{V} \quad (3.2)$$

Ekuilibrum pasar dengan asumsi pasar yang sempurna diperoleh ketika penawaran dan permintaan Bitcoin adalah jumlah yang sama. Oleh karena itu keseimbangan tercapai pada:

$$P_B = \frac{PE}{VB} \quad (3.3)$$

Persamaan ekuilibrium ini menyiratkan bahwa di pasar yang sempurna, harga Bitcoin dalam dolar dipengaruhi secara proporsional oleh tingkat harga barang atau jasa umum dikalikan dengan ukuran ekonomi Bitcoin, dan berbanding terbalik dengan kecepatan Bitcoin dikalikan dengan kapasitas pasar Bitcoin. Tingkat harga barang atau jasa umum,  $P$ , dapat ditentukan secara tidak langsung dari indeks makroekonomi global di pasar yang sebenarnya. Nilai tukar antara beberapa mata uang fiat dan harga Bitcoin menggambarkan hubungan antara pasar yang sebenarnya dan pasar Bitcoin. Perbedaan utama antara pasar Bitcoin dan pasar mata uang umum berasal dari fakta bahwa Bitcoin adalah mata uang virtual berdasarkan teknologi Blockchain. Oleh karena itu, ukuran ekonomi,  $E$ ; kecepatan,  $V$ ; dan kapasitas pasar Bitcoin,  $B$ , terkait erat dengan beberapa variabel pasar terukur yang diambil dari platform Blockchain dan, yang akan menjadi ditinjau dalam subbagian berikutnya (Jang & Lee, 2017).

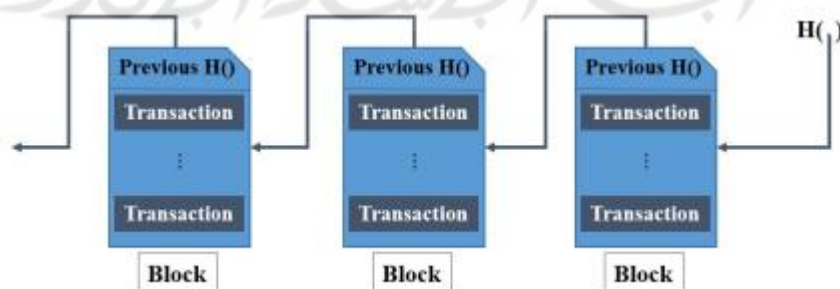
### 3.4 Blockchain

Blockchain merupakan *database* terdistribusi yang menyimpan data catatan yang terus bertambah, dikendalikan oleh beberapa *entity*. Blockchain (*distributed ledger*) adalah sistem layanan yang dapat dipercaya ke sekelompok *node* atau pihak yang tidak saling percaya satu sama lain, umumnya blockchain bertindak sebagai pihak ketiga yang terpercaya dan dapat diandalkan untuk mempertahankan keadaan bersama, menengahi pertukaran, dan menyediakan mesin komputasi yang aman (Cachin & Vukolić, 2017).

Desentralisasi adalah nilai yang dikejar oleh semua mata uang kripto yang bertentangan dengan mata uang fiat umum yang dinilai oleh bank-bank sentral. Desentralisasi dapat ditentukan dengan tujuan berikut: (i) Siapa yang akan

memelihara dan mengelola buku besar transaksi? (ii) Siapa yang berhak untuk memvalidasi transaksi? (iii) Siapa yang akan membuat Bitcoin baru? Blockchain adalah satu-satunya teknologi yang tersedia yang dapat secara bersamaan mencapai tiga tujuan ini. Generasi blok di Blockchain, yang secara langsung terlibat dalam penciptaan dan perdagangan Bitcoin, secara langsung mempengaruhi penawaran dan permintaan Bitcoin. Kombinasi teknologi Blockchain dan pasar Bitcoin adalah contoh dunia nyata dari kombinasi kriptografi tingkat tinggi dan ekonomi pasar.

Kemudian Narayanan et al. (2016) dalam jurnal yang ditulis Jang dan Lee (2017) menjelaskan secara rinci bagaimana Blockchain dapat mencapai tujuan di atas dalam lingkungan Bitcoin. Seorang peserta dalam jaringan Bitcoin bertindak sebagai bagian dari sistem jaringan dengan menyediakan sumber daya perangkat keras dari komputer mereka sendiri, yang disebut “sistem terdistribusi”. Semua penerbitan dan transaksi uang dilakukan melalui jaringan *peer-to-peer*. Semua riwayat perdagangan dicatat dalam Blockchain dan dibagikan oleh jaringan, dan semua riwayat transaksi sebelumnya diverifikasi oleh semua peserta jaringan. Unit ini disebut “blok”, yang mencakup transaksi terbaru dan nilai hash dari blok sebelumnya, membuat data yang tidak dapat diubah dengan fungsi hash, dan ditunjukkan dari blok berikutnya. Gambar 3.2 menunjukkan struktur umum Blockchain. Diperlukan lebih dari waktu tertentu untuk menghasilkan blok sehingga mustahil untuk memalsukan semua atau sebagian dari Blockchain. Algoritma ini disebut *Proof of Work* (PoW), dan kesulitannya secara otomatis diatur untuk memastikan bahwa masalah dapat diselesaikan dalam waktu sekitar 10 menit. PoW juga memberikan insentif untuk memotivasi peserta untuk mempertahankan nilai Bitcoin dengan membayar Bitcoin untuk peserta yang membuat blok.



**Gambar 3.2** Pembentukan Blockchain (Jang & Lee, 2017)

Algoritma kesepakatan PoW dilengkapi dengan beberapa risiko yang melekat. Pertama, validitas blok dapat diintervensi ketika sebagian besar peserta diduduki oleh kelompok dengan tujuan spesifik yang disebut masalah 51%. Kedua, ketika Blockchain bercabang, banyak waktu yang diperlukan untuk membentuk Blockchain yang disepakati sampai rantai terpanjang dipilih setelah beberapa blok dibuat. Kondisi ini menyebabkan penundaan transaksi karena transaksi tidak dapat diselesaikan selama waktu itu. Terakhir, mungkin ada batas kapasitas Blockchain atau batas kinerja setiap node. Keamanan Blockchain saat ini dapat dipantau dengan mengamati variabel terukur dalam Blockchain dari <https://blockchain.info/>.

Ada banyak variabel yang ditulis di buku besar publik seperti (Blockchain, 2017):

1. *Average Block Size* (Rata-rata Ukuran Blok) : ukuran rata-rata blok dalam MB.
2. *Transactions per Block* (Transaksi per Blok) : jumlah rata-rata transaksi per block.
3. *Median Transaction Confirmation Time (with fee)* (Median Konfirmasi Transaksi Waktu (dengan biaya)) : waktu rata-rata untuk transaksi yang akan diterima ke blok tambang.
4. *Hash Rate* (Tingkat Hash) : perkiraan jumlah hash tera per detik jaringan Bitcoin kinerja.
5. *Difficulty* (Kesulitan) : ukuran relatif dari betapa sulitnya untuk menemukan blok baru.
6. *Cost% of Transaction Volume* (Biaya% Transaksi Volume) : sebuah grafik yang menunjukkan penambang pendapatan sebagai persentase dari volume transaksi.
7. *Mining Revenue* (Penghasilan Pertambangan) : jumlah total hadiah blok coinbase dan ongkos-ongkos transaksi yang dibayarkan kepada para penambang.
8. *Unique Addresses* (Alamat Unik) : jumlah total alamat unik yang digunakan pada blockchain Bitcoin.

9. *Total Number of Transactions per Day* (Total Jumlah Transaksi per Hari) : jumlah transaksi Bitcoin harian dikonfirmasi.

### **3.5 Indeks Harga Saham**

Indeks harga saham adalah suatu indikator yang menunjukkan pergerakan harga saham dalam suatu periode. Indeks ini berfungsi sebagai indikator trend pasar, yang berarti pergerakan indeks menggambarkan kondisi pasar pada suatu saat, apakah keadaan pasar sedang aktif atau sedang lesu. Dengan adanya indeks, dapat diketahui trend pergerakan harga saham saat ini; apakah sedang naik, stabil atau turun. Misal jika diawal bulan nilai indeks 300 dan saat ini di akhir bulan menjadi 360, maka dapat dikatakan bahwa secara rata-rata harga saham mengalami peningkatan sebesar 20%.

Pergerakan indeks menjadi indikator penting bagi para investor untuk menentukan apakah mereka akan menjual, menahan atau membeli suatu atau beberapa saham. Karena harga-harga saham bergerak dalam hitungan detik dan menit, maka nilai indeks pun bergerak turun naik dalam hitungan waktu yang cepat (Infovesta, 2019).

### **3.6 Standard&Poor's 500 (S&P 500)**

S&P 500 adalah salah satu indeks yang menjadi indikator utama dari pergerakan pasar saham Amerika Serikat. Isinya 505 saham yang diterbitkan oleh 500 emiten besar. Kapitalisasi emiten yang masuk sebagai konstituen pada indeks ini minimal senilai 6,1 miliar dolar Amerika Serikat (AS). Indeks ini merupakan indikator untuk mencerminkan kinerja saham-saham dari perusahaan berkapitalisasi besar.

Indeks ini diperkenalkan oleh lembaga pemeringkat Standard & Poor's pada 1 Januari 1957 diposisi 386. Dari tahun 1969 hingga awal tahun 1981, indeks S&P secara bertahap turun, menjadi di bawah 300. Indeks S&P 500 lebih disukai dipakai sebagai ukuran saham di AS karena anggotanya yang banyak. Indeks S&P menggunakan metodologi kapitalisasi pasar, sehingga membuat bobot perusahaan berkapitalisasi besar lebih banyak (Candra, 2018).

### 3.7 Euro STOXX 50

Indeks pasar saham EURO STOXX 50 adalah representasi 50 perusahaan blue-chip terbesar di Eropa yang secara aktif beroperasi di Zona Euro, dengan kapitalisasi pasar. Indeks ini mencakup saham-saham dari 11 negara di Zona Euro, yaitu Austria, Belgia, Irlandia, Italia, Belanda, Luksemburg, Spanyol, Portugal, dan Finlandia, dengan Prancis dan Jerman sebagai pemain utamanya.

EuroStoxx 50 dikelola dan dilisensikan oleh STOXX Limited, yang merupakan perusahaan gabungan dari Deutsche Boerse AG, Dow Jones & Company, dan SWX Group. Stoxx Ltd. merupakan penyedia indeks yang terintegrasi secara global yang mengindikasikan investasi pasar ekuitas di seluruh dunia. Agar dapat bergabung dalam indeks Euro STOXX 50, suatu negara harus menjadi negara anggota Zona Euro (Avatrade, 2018a).

### 3.8 Dow 30

Dow Jones Industrial Average (DJI) atau biasa disebut the Dow/Dow 30 adalah sebuah indeks yang terdiri dari 30 perusahaan besar yang ada dan listing di bursa saham AS. Nilai dari Dow Jones diukur secara aritmatika yang berarti nilainya bukan berdasarkan kapitalisasi dari perusahaan yang ada didalamnya melainkan dengan harga dari sebuah saham dari masing-masing perusahaan yang ada didalamnya. Nilai ini bisa dikoreksi sesuai dengan kondisi yang ada, misalnya saja perusahaan melakukan stock split maka hal tersebut dapat menurunkan nilai per lembar sahamnya dan menyebabkan indeks menjadi turun drastis, hal ini dikoreksi sehingga indeks Dow masih tetap stabil dan relevan terhadap perubahan diluar harga yang asli.

*Dow Jones Industrial Average* merupakan salah satu indeks tertua yang ada di AS. Dulu di bursa saham tidak terdapat indeks sehingga investor tidak tahu apakah pasar sedang bagus atau turun. Investor juga tidak dapat membandingkan kinerja portfolionya dengan kinerja pasar, karena memang belum ada nilainya. Charles Dow berpikir demikian, sehingga dia membuat sebuah indeks untuk menggambarkan keadaan sebenarnya yang terjadi di pasar. Pada 26 Mei 1896 dia



mempublikasikan sebuah indeks yang bernama *Dow Jones Industrial Average* (DJI) yang menjadi dasar pada Dow Jones sekarang ini (Herlambang, 2018).

### **3.9 Indeks Nasdaq**

Indeks Nasdaq adalah indeks pada bursa kedua terbesar setelah bursa New York yaitu bursa Nasdaq. Baik bursa New York maupun Nasdaq terletak di New York City. Bursa Nasdaq mulai beroperasi pada tahun 1971 dan merupakan bursa elektronik pertama waktu itu. Indeks utamanya adalah Nasdaq Composite yang berisi sekitar 3.000 saham yang tercatat di bursa Nasdaq. Emiten yang masuk ke indeks Nasdaq terdiri atas berbagai macam perusahaan pada berbagai macam sektor, kecuali sektor keuangan. Indeks ini disusun berdasarkan metode kapitalisasi, bobot setiap saham pada indeks tergantung dari kapitalisasinya. Setiap kuartal ada penyesuaian saham-saham dalam indeks ini (Candra, 2018).

### **3.10 Crude Oil**

Minyak mentah atau *crude oil* adalah salah satu sumber daya yang sangat vital saat ini, dikarenakan hasil olahan minyak mentah merupakan sumber energi. Minyak mentah dapat diolah menjadi sumber energi, seperti *Liquified Petroleum Gas* (LPG), bensin, solar, minyak pelumas, minyak bakar dan lain-lain. Harga minyak dunia diukur menggunakan harga spot pasar minyak dunia. Standar harga minyak dunia umumnya mengacu pada *West Texas Intermediate* (WTI). WTI merupakan minyak bumi berkualitas tinggi yang diproduksi di Amerika Utara (SeputarForex, 2016). Minyak mentah sama halnya dengan mata uang dan emas yang merupakan salah satu indikator yang terlibat dalam ekonomi dunia, dikarenakan volatilitas mengikuti peristiwa ekonomi dan politik suatu negara.

Harga Minyak Dunia yang berfluktuatif dapat mempengaruhi pasar modal suatu negara. Pada umumnya kenaikan harga minyak akan mendorong kenaikan harga saham sektor pertambangan, karena dengan peningkatan harga minyak akan memicu kenaikan harga bahan tambang lainnya. Hal tersebut akan mengakibatkan perusahaan pertambangan berpotensi untuk meningkatkan labanya (Sartika, 2017).

### 3.11 Gold Futures

Emas merupakan salah satu jenis komoditi yang paling banyak diminati untuk tujuan investasi. Di samping itu, emas juga digunakan sebagai standar keuangan atau ekonomi, cadangan devisa dan alat pembayaran yang paling utama di beberapa negara. Para investor umumnya membeli emas untuk *hedge* atau *safe haven* terhadap beberapa krisis termasuk ekonomi, politik, sosial atau krisis yang berbasis mata uang (Futures, 2018).

Harga emas yang dijadikan patokan di seluruh dunia sejak tahun 1968 adalah harga emas berdasarkan standar pasar emas London. Sistem ini dinamakan London *Gold Fixing*. Menurut Sunariyah Emas merupakan salah satu bentuk investasi yang cenderung bebas resiko. Emas memiliki nilai yang cenderung stabil dan naik. Sangat jarang sekali harga emas turun. Emas merupakan alat yang dapat digunakan untuk menangkal inflasi yang kerap terjadi setiap tahunnya (Sartika, 2017).

### 3.12 Volatility Index (VIX)

Indeks Volatilitas CBOE (VIX) secara luas dianggap sebagai indikator utama volatilitas dan sentimen investor pasar saham. Ini adalah ukuran ekspektasi pasar akan volatilitas jangka pendek dari harga opsi indeks saham S&P 500. Chicago Board Options Exchange (CBOE) memperkenalkan VIX pada tahun 1993 untuk mengukur ekspektasi pasar terhadap volatilitas 30 hari yang tersirat dari opsi S&P 100 *at-the-money* di mana harga kesepakatan sama dengan harga sekuritas acuan. Sejak saat itu, indeks tumbuh menjadi standar untuk mengukur volatilitas pasar dalam pasar bursa AS. Hal ini memberikannya julukan ‘indeks ketakutan’ dan ‘barometer ketakutan’.

Pada tahun 2003, didorong oleh tingkat kepentingan indeks yang terus bertambah, CBOE dan Goldman Sachs memperbarui VIX untuk mencerminkan status patokannya. VIX sekarang berdasarkan indeks yang lebih luas, S&P 500, yang memungkinkan gambaran yang lebih akurat mengenai volatilitas pasar yang diharapkan (Avatrade, 2018c).

### 3.13 Nikkei 225

Indeks Nikkei 225 merupakan indeks utama di bursa Tokyo. Indeks Nikkei dipandang sebagai salah satu barometer penting perekonomian dan pasar saham Jepang. Indeks Nikkei disebut-sebut sama seperti indeks Dow Jones. Bahkan, dari tahun 1975 hingga 1985, indeks ini sempat disebut *Nikkei Dow Jones Stock Average*.

Indeks ini merupakan indeks tertua di kawasan Asia mulai diperkenalkan pada tahun 1950. Sama seperti indeks Dow Jones yang dibuat oleh surat kabar, indeks Nikkei juga diciptakan oleh surat kabar ekonomi Jepang Nihon Keizai Shimbun. Perhitungan indeks Nikkei juga mengikuti harga, sama seperti Dow Jones. Konstituen indeks Nikkei dihitung dan dikaji setiap September (Candra, 2018).

### 3.14 FTSE 100

FTSE 100 adalah indeks saham dari 100 perusahaan yang terdaftar di LSE, yang diukur berdasarkan kapitalisasi pasar, dan dikelola oleh grup FTSE. Saat pasar dibuka, indeks ini diperbarui dan diterbitkan setiap 15 detik. Indeks ini memasuki pasar pada 3 Januari 1984 pada level bawah sebesar 1.000 poin. Di era berjayanya dot-com, pada tahun 2000 indeks tersebut mencapai level 7.103,98 poin. Setelah itu, indeks ini mengalami penurunan selama krisis keuangan 2007-2010 ke level rendah berada di bawah 3.500. Tujuh tahun kemudian pada Maret 2017, nilai intraday tertinggi yang dicapai dan tercatat adalah 7.777,62 (Avatrade, 2018b).

### 3.15 Jakarta Composite Index (JKSE)

Indeks harga saham gabungan merupakan indeks yang menggunakan semua emiten di Bursa Efek Indonesia (BEI) untuk dijadikan sebagai komponen dalam perhitungan indeks harga. Di samping itu, Bursa Efek Indonesia memiliki hak untuk memasukkan dan mengeluarkan beberapa emiten/perusahaan yang tercatat supaya bisa menunjukkan keadaan di pasar modal yang wajar.

Indeks saham IHSG atau biasanya disebut sebagai Jakarta Composite Index (JKSE) ini dibentuk pertama kali pada tanggal 1 April 1983 sebagai indikator adanya pergerakan dan perubahan harga saham di BEI. Namun, kenaikan dan

penurunan IHSG bukan mengindikasikan bahwa semua harga saham di BEI mengalami kenaikan atau penurunan. Alasannya adalah karena apabila IHSG naik, maka sebagian besar harga saham yang tercatat di bursa cenderung naik daripada harga saham yang sedang mengalami penurunan dan sebaliknya (Utari, 2016).

### 3.16 Indeks LQ45

Jenis indeks saham di BEI selanjutnya adalah indeks saham LQ45. Jenis indeks saham ini adalah indeks saham alternatif selain IHSG. Selain sebagai pelengkap IHSG, tujuan dibuatnya indeks LQ45 supaya dapat dijadikan sarana yang obyektif, terpercaya oleh analis keuangan, manajer investasi dan investor dalam mengawasi pergerakan harga saham-saham di BEI.

Komponen indeks LQ45 adalah 45 emiten yang memiliki likuiditas tinggi dan memiliki kualifikasi berikut (Utari, 2016):

1. Emiten harus memiliki kapitalisasi pasar yang termasuk ke dalam 60 saham dengan nilai kapitalisasi pasar terbesar Di BEI dalam kurun waktu satu tahun.
2. Emiten dalam indeks LQ45 merupakan salah satu 60 saham yang dinilai paling likuid di bursa karena memiliki nilai transaksi perdagangan terbesar di pasar regular.
3. Emiten sudah melakukan listing di Bursa Efek Indonesia minimal tiga bulan terakhir. Oleh karena itu, saham-saham baru yang baru saja melakukan IPO (Initial Public Offer) tidak dapat tergolong dalam indeks LQ45.
4. Emiten mempunyai kinerja dan prospek ke depan yang baik.

### 3.17 GBP/USD (British Pound/U.S. Dollar)

GBP/USD adalah singkatan untuk pasangan mata uang British pound dan U.S. dollar atau biasa disebut *cable*. Pasangan mata uang ini menginformasikan berapa banyak dolar AS (mata uang dikutip) yang diperlukan untuk membeli satu pound Inggris (mata uang dasar). Nilai GBP/USD dikutip sebagai 1 pound Inggris per X dolar AS. Misalnya, jika pasangan ini diperdagangkan pada 1,50, berarti dibutuhkan 1,5 dolar AS untuk membeli 1 pound Inggris (AccelaInfinia, 2019).

### **3.18 JPY/USD (Japanes Yen/U.S. Dollar)**

JPY/USD adalah singkatan untuk pasangan mata uang Japanes Yen dan U.S. dollar. Pasangan mata uang ini menginformasikan berapa banyak dolar AS (mata uang dikutip) yang diperlukan untuk membeli satu Yen Jepang (mata uang dasar). Nilai JPY/USD dikutip sebagai 1 Yen Jepang per X dolar AS. Misalnya, jika pasangan ini diperdagangkan pada 0,008933 berarti dibutuhkan 0,008933 dolar AS untuk membeli 1 Yen Jepang (ValutaFX, 2019b).

### **3.19 CHF/USD (Swiss Franc/U.S. Dollar)**

CHF/USD adalah singkatan untuk pasangan mata uang *Swiss Franc* dan U.S. dollar. Pasangan mata uang ini menginformasikan berapa banyak dolar AS (mata uang dikutip) yang diperlukan untuk membeli satu *Swiss Franc* (mata uang dasar). Nilai CHF/USD dikutip sebagai 1 Swiss Franc per X dolar AS. Misalnya, jika pasangan ini diperdagangkan pada 0,982053, berarti dibutuhkan 0,982053 dolar AS untuk membeli 1 Swiss Franc (TransferWise, 2019).

### **3.20 CNY/USD ( Yuan Cina/U.S. Dollar)**

CNY/USD merupakan singkatan untuk pasangan mata uang Yuan Cina dan U.S. dollar. Pasangan mata uang ini menginformasikan berapa banyak dolar AS (mata uang dikutip) yang diperlukan untuk membeli satu Yuan Cina (mata uang dasar). Nilai CNY/USD dikutip sebagai 1 Yuan Cina per X dolar AS. Misalnya, jika pasangan ini diperdagangkan pada 0,14874, berarti dibutuhkan 0,14874 dolar AS untuk membeli 1 Yuan Cina (ValutaFX, 2019a).

### **3.21 EUR/USD (Euro/U.S. Dollar)**

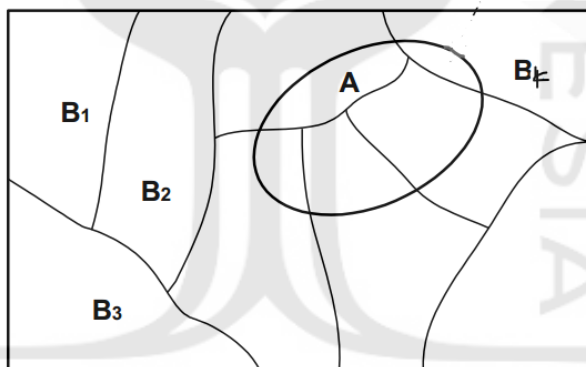
EUR/USD (atau Euro Dollar) pasangan mata uang termasuk dalam kelompok 'Mata uang utama', sebuah cara untuk menyebutkan pasangan terpenting di dunia. Popularitas Euro Dollar disebabkan oleh fakta bahwa ia mengumpulkan dua ekonomi utama: Eropa dan AS. Ini adalah pasangan mata uang yang diperdagangkan secara luas di mana Euro adalah mata uang dasar dan Dolar AS adalah mata uang kontra. Karena EUR/USD terdiri lebih dari separuh volume perdagangan di seluruh dunia di Pasar Forex, hampir tidak mungkin ada gap untuk

muncul, apalagi gap perpecahan konsekuen ke arah yang berlawanan. EUR/USD mencapai tinggi sepanjang waktu 1,87 pada Juli 1973 dan rekor rendah 0,70 pada Februari 1985 (Fxstreet, 2019).

### 3.22 Teorema Bayes

Nama teorema Bayes diambil dari nama penemu teorema tersebut, yaitu Reverend Thomas Bayes (1702-1761). Teorema Bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa, berdasarkan pengaruh yang didapat dari hasil observasi peristiwa sebelumnya. Teorema Bayes menyempurnakan teorema probabilitas bersyarat yang hanya dibatasi oleh dua buah kejadian sehingga dapat diperluas untuk  $k$  buah kejadian. Aplikasi teorema Bayes banyak ditemukan pada bidang komputer cerdas sebagai salah satu dasar dari metode *machine learning* dan data mining.

Diagram Venn Teorema Bayes digunakan ketika ingin menghitung probabilitas  $\rightarrow P(B_1|A), P(B_2|A), \dots, P(B_k|A)$ .



**Gambar 3.3** Diagram Venn Teorema Bayes (Walpole & Myers, 1995)

Misalkan kejadian  $\{B_1, B_2, \dots, B_k\}$  merupakan suatu sekatan (partisi) dari ruang sampel S dengan  $P(B_i) \neq 0$  untuk  $i = 1, 2, \dots, k$ . Dan misalkan A suatu kejadian sembarang dalam S dengan  $P(A) \neq 0$ , untuk  $k=3$ , maka setiap kejadian A anggota S adalah:

$$P(A) = \sum_{i=1}^k P(B_i \cap A) = \sum_{i=1}^k P(B_i)P(A|B_i) \quad (3.4)$$

Berdasarkan teorema Probabilitas Bersyarat: Probabilitas bersyarat suatu peristiwa A, dengan syarat peristiwa B didefinisikan sebagai:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.5)$$

Atau

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad (3.6)$$

Dimana berdasarkan teori himpunan diketahui:

$$P(A \cap B) = P(B \cap A) \quad (3.7)$$

Sehingga dari Persamaan 3.6 dengan Persamaan 3.7 didapatkan:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad (3.8)$$

Maka

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(B|A) &= \frac{P(A|B)P(B)}{P(A)} \end{aligned} \quad (3.9)$$

Berdasarkan hubungan probabilitas A dengan probabilitas kejadian bersyarat sebagaimana ditunjukkan persamaan (3.4), yaitu:

$$P(A) = \sum_i^k P(A|B_i)P(B_i) \quad (3.10)$$

Sehingga probabilitas suatu kejadian yang dibatasi oleh  $k$  buah kejadian sebagai syaratnya akan diperoleh dari penurunan rumus sebagai berikut:

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^k P(B_i)P(A|B_i)}; i = 1, 2, \dots, k \quad (3.11)$$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_k)P(B_k)} \quad (3.12)$$

Secara umum formula teorema Bayes adalah sebagai berikut (Walpole & Myers, 1995):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.13)$$

Dimana:

- Peluang Posterior  $P(A|B)$  adalah prediksi peluang munculnya satu kejadian berdasarkan informasi dari kejadian yang lain.
- Peluang Prior  $P(A)$  adalah peluang munculnya suatu kejadian yang sudah kita yakini sebelumnya dan bisa jadi kejadian ini dipengaruhi kejadian yang lain.

- Peluang Likelihood  $P(B|A)$  adalah peluang yang menyatakan derajat kemungkinan pengaruh suatu informasi kejadian terhadap kejadian yang lain.
- Peluang *Evidence*  $P(B)$  adalah sebuah ukuran pembanding konstan berdasarkan peluang suatu informasi kejadian.

### 3.23 Imputasi Nilai Hilang dengan Rata-rata Bergerak Tertimbang

Penggantian nilai hilang dengan rata-rata bergerak tertimbang. Menggunakan ukuran jendela semi adaptif untuk memastikan semua NA diganti. Berikut adalah sintaksnya (Moritz & Bartz-Beielstein, 2017):

```
na.ma(x, k = 4, weighting = "exponential")
```

**Gambar 3.4** Sintaks Imputasi Nilang Hilang

Dimana:

$x$  : Vektor Numerik (vektor) atau Seri Waktu (ts) objek di mana nilai-nilai yang hilang harus diganti.

$k$  : Lebar integer dari jendela rata-rata bergerak. Perluas ke kedua sisi elemen tengah mis.  $k = 2$  berarti 4 pengamatan (2 kiri, 2 kanan) diperhitungkan. Jika semua pengamatan di jendela saat ini adalah NA, ukuran jendela secara otomatis meningkat hingga setidaknya ada 2 nilai non-NA.

*weighting* : Bobot untuk digunakan. Terima input berikut:

- "simple" - *Simple Moving Average* (SMA)
- "linear" - *Linear Weighted Moving Average* (LWMA)
- "eksponensial" - *Average Moving Average Weighted* (EWMA)

Dalam fungsi ini nilai-nilai yang hilang diganti dengan memindahkan nilai rata-rata. *Moving Averages* juga kadang-kadang disebut sebagai "*moving average*", "*rolling average*", "*rolling average*" atau "*running average*".

*Mean* dalam implementasi ini diambil dari jumlah pengamatan yang sama di kedua sisi dari nilai pusat. Ini berarti untuk nilai NA pada posisi  $i$  dari deret waktu, pengamatan  $i-1$ ,  $i+1$  dan  $i+1$ ,  $i+2$  (dengan asumsi ukuran jendela  $k = 2$ ) digunakan untuk menghitung rata-rata. Karena bisa dalam kasus kesenjangan NA panjang juga terjadi, bahwa semua nilai di samping nilai pusat juga NA, algoritma ini memiliki



ukuran jendela semi-adaptif. Setiap kali ada kurang dari 2 nilai non-NA di jendela lengkap yang tersedia, ukuran jendela meningkat secara bertahap, sampai setidaknya 2 nilai non-NA ada di sana. Dalam semua kasus lain, algoritma tersebut menempel ukuran jendela yang telah ditetapkan.

Ada beberapa pilihan untuk menggunakan SMA, LWMA dan EWMA.

SMA : semua pengamatan di jendela diberi bobot yang sama untuk menghitung rata-rata.

LWMA : penurunan bobot dalam perkembangan aritmatika. Pengamatan langsung di sebelah nilai sentral  $i$ , memiliki bobot  $1/2$ , pengamatan satu lagi lebih jauh ( $i-2, i + 2$ ) memiliki bobot  $1/3$ , selanjutnya ( $i-3, i + 3$ ) memiliki bobot  $1/4, \dots$

EWMA : menggunakan faktor bobot yang menurun secara eksponensial. Pengamatan langsung di sebelah nilai sentral  $i$ , memiliki bobot  $1/2^1$ , pengamatan satu lebih jauh ( $i-2, i + 2$ ) memiliki bobot  $1/2^2$ , selanjutnya ( $i-3, i + 3$ ) memiliki berat  $1/2^3, \dots$

### 3.24 *Machine Learning*

Secara definisi, *machine learning* adalah cabang dari ilmu kecerdasan buatan yang berfokus pada pembangunan dan studi sebuah sistem agar mampu belajar dari data-data yang diperolehnya. Menurut Arthur Samuel, *machine learning* adalah bidang studi yang memberikan kemampuan program komputer untuk belajar tanpa secara eksplisit diprogram.

Untuk bisa mengaplikasikan teknik-teknik *machine learning* maka harus ada data. Tanpa data maka algoritma *machine learning* tidak dapat bekerja. Data yang ada biasanya dibagi menjadi dua, yaitu data *training* dan data *testing*. Data training digunakan untuk melatih algoritma, sedangkan data testing digunakan untuk mengetahui performa algoritma yang telah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat (Fikriya, Irawan, & Soetrisno, 2017).

*Machine Learning* merupakan rangkaian teknik yang mampu membantu untuk menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data besar tersebut menggunakan algoritma pembelajaran.

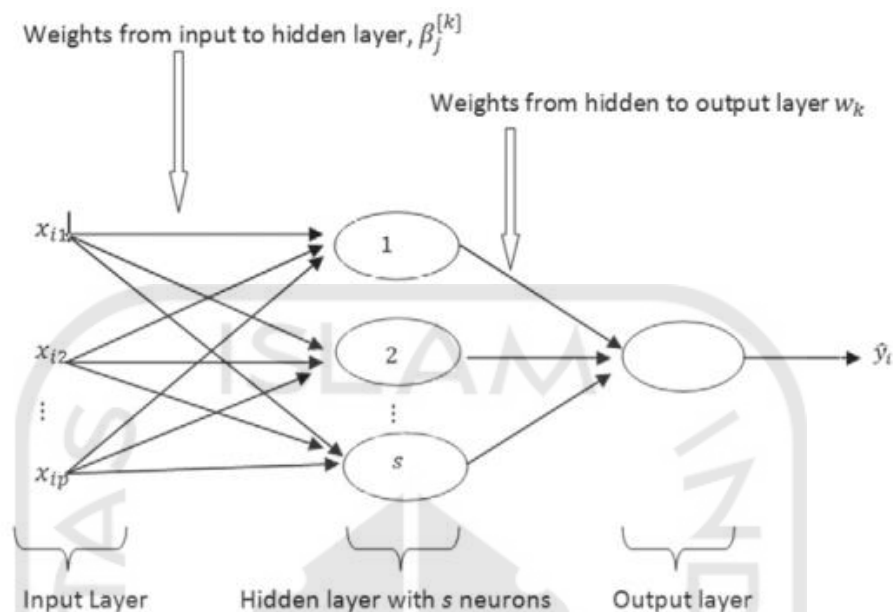
*Machine Learning* dapat membuat komputer memprogram diri mereka sendiri. Pada dasarnya, *Machine Learning* membiarkan data untuk melakukan pekerjaannya sendiri (Danukusumo, 2017).

### 3.24.1 *Neural Network*

*Neural Network* (NN) adalah prosesor yang didistribusikan secara masif yang terdiri dari elemen pemrosesan sederhana yang disebut neuron, yang memodelkan beberapa fungsi seperti otak manusia (Haykin, 1998). Penggunaan NN menawarkan beberapa sifat dan kemampuan yang berguna: - Nonlinier, Adaptivitas, Respons Evidensial, Keyakinan dalam pengambilan keputusan. Keuntungan utama dari sistem pembelajaran adalah bahwa mereka nonparametrik; model prediksi dapat disesuaikan dengan data di situs tertentu (Srinivasan & Fisher, 1995).

Salah satu NN yang paling dasar dan terkenal adalah *Single Hidden Layer Feed Forward Neural Network* (SLNN), juga dikenal sebagai perceptron, ditunjukkan pada Gambar 3.5. Prediksi dari kelas jaring ini dapat diperoleh dalam 2 langkah: pertama *input* secara nonlinier ditransformasikan dalam lapisan tersembunyi, dan kemudian *output* dari lapisan tersembunyi digabungkan secara linier untuk mendapatkan prediksi, seperti yang diberikan oleh (Gianola, Okut, Weigel, & Rosa, 2011). Modelnya kemudian:

$$y_i = \mu + \sum_{k=1}^s w_k g_k \left( b_k + \sum_{j=1}^p x_{ij} \beta_j^{[k]} \right) + \varepsilon_i \quad (3.14)$$



**Gambar 3.5** Ilustrasi Jaringan Umpan Maju Neural *single layer* (Gianola et al., 2013).

Dimana  $(w_1, \dots, w_s)$  adalah bobot jaringan;  $(b_1, \dots, b_s)$  disebut bias dalam pembelajaran mesin;  $(\beta_1^{[1]}, \dots, \beta_p^{[1]}; \dots; (\beta_1^{[s]}, \dots, \beta_p^{[s]})$  adalah kekuatan koneksi, di mana  $(\beta_j^{[k]})$  menunjukkan parameter untuk *input*  $j$  di neuron  $k = 1, \dots, s$ , dan  $g_k(\cdot)$  adalah fungsi aktivasi, yaitu, fungsi yang memetakan *input* dari garis nyata ke dalam interval terbuka yang dibatasi  $(-1,1)$ . Sebagai contoh,  $g_k(x) = 2/[1 + \exp(-2x)] - 1$  dikenal sebagai fungsi aktivasi tanh (*tangent hyperbolic*).

### 3.2.4.2 Bayesian Regularized Neural Networks

SLNN yang dijelaskan di atas cukup fleksibel untuk mendekati fungsi nonlinier. Ketika jumlah *input* ( $p$ ) dan jumlah neuron ( $s$ ) meningkat, jumlah parameter untuk diperkirakan dalam jaringan juga meningkat. Bersamaan dengan ini, fleksibilitas jaringan saraf dapat menyebabkan masalah serius *overfitting*. Untuk menghindari masalah ini, biasanya digunakan metode estimasi hukuman yang diterapkan, misalnya, menggunakan pendekatan Bayesian. Penyusutan mengarah ke beberapa distribusi sebelumnya. MacKay (1994) mengembangkan algoritma yang digunakan untuk memperoleh estimasi semua parameter dalam BRNN

menggunakan pendekatan *Empirical Bayes* (EB). Deskripsi singkat disediakan di bawah ini (Mackay, 1992) .

Misalkan  $\theta = (w_1, \dots, w_s; b_1, \dots, b_s; \beta_1^{[1]}, \dots; \beta_p^{[1]}, \dots; \beta_1^{[s]}, \dots; \beta_p^{[s]})$  menjadi vektor bobot, bias, dan kekuatan koneksi. Perhatikan bahwa  $\mu$  tidak termasuk dalam  $\theta$ , karena parameter ini dapat dengan mudah dihilangkan, misalnya, cukup dengan memusatkan vektor respons. Misalkan  $p(\theta|\sigma_\theta^2) = MN(0, \sigma_\theta^2 I)$  menjadi distribusi sebelumnya, di mana MN adalah distribusi normal multivariat, dan  $\sigma_\theta^2$  adalah varian yang umum untuk semua elemen  $\theta$ . Dalam pendekatan EB, 2 langkah ini diulangi secara berulang sampai konvergen:

1. Dapatkan mode posterior bersyarat dari elemen dalam  $\theta$ , dengan asumsi  $\sigma_e^2$  dan  $\sigma_\theta^2$  diketahui. Mode-mode ini diperoleh dengan memaksimalkan

$$p(\theta|y, \sigma_e^2, \sigma_\theta^2) = \frac{p(\theta|y, \sigma_e^2)p(\theta|\sigma_\theta^2)}{p(y|\sigma_e^2, \sigma_\theta^2)} \quad (3.15)$$

yang setara dengan meminimalkan jumlah kuadrat yang diperluas

$$F(\theta) = \frac{1}{2\sigma_e^2} \sum_{i=1}^n e_i^2 + \frac{1}{2\sigma_\theta^2} \sum_{j=1}^m \theta_j^2 \quad (3.16)$$

Dimana  $e_i = y_i - \hat{y}_i$

2. Perbarui komponen varians  $\sigma_e^2$  dan  $\sigma_\theta^2$  dengan memaksimalkan  $p(y|\sigma_e^2, \sigma_\theta^2)$ . Karena nonlinier, *log-likelihood* marjinal,  $\log p(y|\sigma_e^2, \sigma_\theta^2)$  tidak dapat diekspresikan dalam bentuk tertutup, tetapi dapat diperkirakan sebagai

$$G(\alpha, \beta) = \log p(y|\sigma_e^2, \sigma_\theta^2) \quad (3.17)$$

$$\approx c + \frac{n}{2} \log \beta + \frac{m}{2} \log \alpha - \frac{1}{2} \log |\Sigma|_{\theta=\theta^{map}}$$

Dimana  $c$  adalah konstanta;  $\Sigma = \nabla^2 F(\theta)$  adalah matriks turunan kedua dari orde  $m \times m$ ;  $m$  adalah urutan  $\theta$ , dan peta berarti maksimum a posterior.

Gianola et al. (2011) dan (Fisher, 1918) dalam jurnal (Gianola et al., 2013) telah menunjukkan bahwa model aditif standar sangat kecil dari genetika kuantitatif dapat dilihat sebagai NN dengan neuron tunggal. Ini dapat dituliskan sebagai berikut:

$$y = u + \varepsilon \quad (3.18)$$

dimana  $u \sim MN(0, \sigma_u^2 A)$  adalah vektor efek aditif,  $A$  adalah matriks hubungan aditif yang dihitung dari silsilah, dan  $\varepsilon \sim MN(0, \sigma_e^2 I)$ . Model diberikan dalam persamaan (3.18) dapat ditulis ulang sebagai:

$$y = u + \varepsilon = Lz\sigma_u + \varepsilon = Lu^* + \varepsilon \quad (3.19)$$

dimana  $L = \{l_{ij}\}$  adalah matriks segitiga bawah yang diperoleh dari dekomposisi Cholesky  $A$ , yaitu,  $LL' = A$ ,  $z \sim MN(0, I)$ , dan  $u^* = z\sigma_u \sim MN(0, \sigma_u^2 I)$ . Jelas bahwa persamaan (3.16) adalah kasus khusus persamaan (3.14) diperoleh dengan mengambil  $\mu = 0$  ( $y$  dipusatkan pada  $0$ ),  $s = 1$ ,  $w_1 = 1$ ,  $b_1 = 0$ ,  $x_{ij} = l_{ij}$  dan  $\beta_j^{[1]} = u_j^*$ ,  $i, j = 1, \dots, n$  dengan  $g_1$  menjadi operator identitas. Gianola et al. (2011) menunjukkan bahwa kemampuan prediktif model ini dapat berpotensi ditingkatkan dengan memasukkan neuron selama mempengaruhi transformasi nonlinear, yang mengarah pada model yang diberikan dalam persamaan (3.14). Secara teori, ini memberikan model fleksibilitas yang besar, karena informasi hubungan aditif digunakan dengan mahir.

Ketika jumlah prediktor meningkat, demikian juga jumlah parameter untuk diperkirakan. Karena itu, ketika berhadapan dengan data genotipik dimensi tinggi, disarankan untuk mengurangi dimensi masalah. Dalam kasus NN, ini mudah karena seperti ditunjukkan di atas, model infinitesimal dapat dilihat sebagai regresi pada silsilah atau pada matriks hubungan genomik, seperti yang ditunjukkan oleh de los Campos et al. (2009) dan Gianola et al. (2011) dalam jurnal Gianola et al. (2013). Misalnya, matriks hubungan genom yang disarankan oleh VanRaden (2008) dalam jurnal Gianola et al. (2013) sebagai: misalnya, matriks hubungan genom:

$$G = \frac{XX'}{2 \sum_{j=1}^p p_j(1 - p_j)} \quad (3.20)$$

dimana  $X = \{x_{ij}\} \in \{-1, 0, 1\}$  adalah matriks kode SNP dimensi  $n \times p$  dan  $p_j$  adalah frekuensi alel minor untuk SNP  $j$ ,  $j = 1, \dots, p$ . Singkatnya, NN dasar yang diberikan oleh persamaan (3.14) dapat dipertegas sebagai prediktor: i) matriks kejadian  $X$ , ii) informasi silsilah (matriks  $A$ ), atau iii) matriks hubungan genomik (mis.,  $G$ ).

Biasanya, akan lebih mudah untuk menskalakan *input* untuk berada di  $[-1, 1]$ , karena perilaku NN sensitif terhadap skala. Dalam pendekatan yang digunakan di sini, data pelatihan disusun kembali untuk berada di  $[-1, 1]$ , informasi normalisasi dipertahankan sehingga begitu model itu hilang, prediksi dapat diberikan dalam skala asli. Informasi *input* untuk set pengujian (jika ada) dihitung ulang menggunakan informasi normalisasi yang diperoleh dari set pelatihan, maka prediksi dibuat dan *output* dinaikkan sehingga dipetakan kembali ke skala asli. Pendekatan ini digunakan misalnya dalam Matlab R2010b (lihat fungsi `premnmx`, `postmnmx`, dan `trmnmx`). Gianola et al. (2011) mengilustrasikan dengan contoh sederhana bagaimana penyelamatan dapat dilakukan.

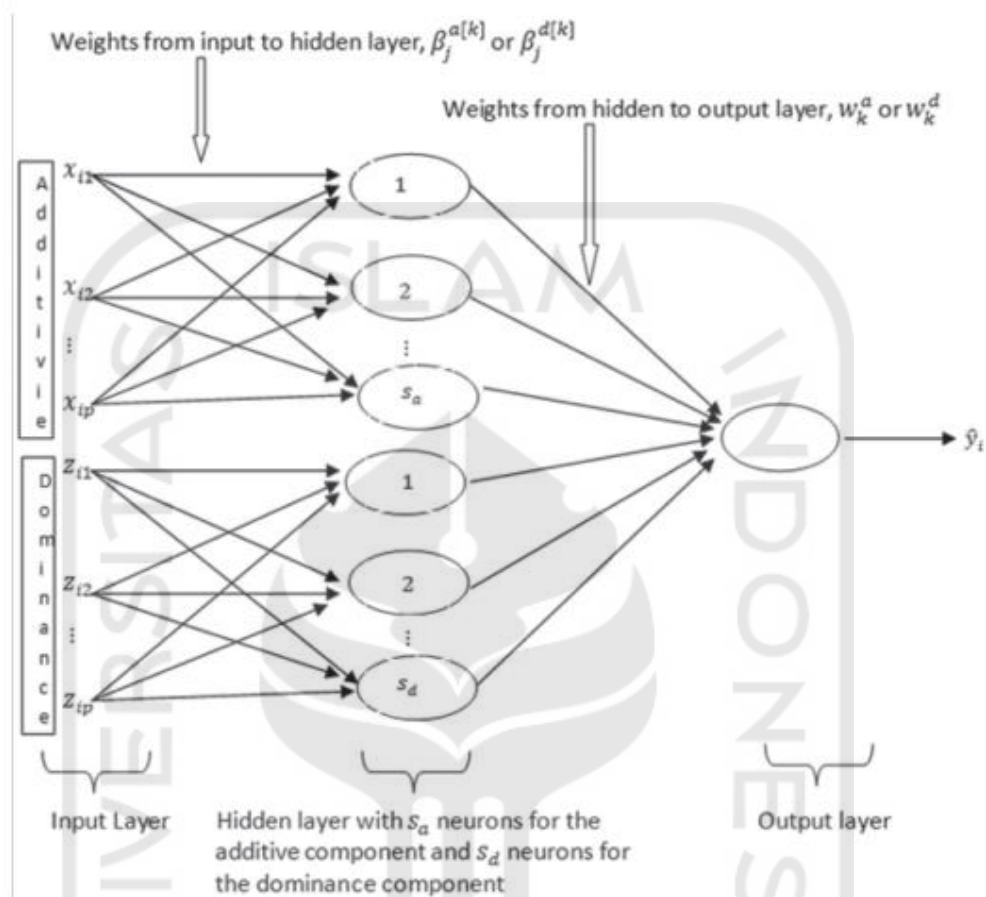
### 3.24.3 Jaringan Saraf Tiruan dengan Efek Aditif dan Dominasi

Sebagian besar model yang digunakan dalam seleksi genom berfokus pada efek aditif saja (Gianola et al., 2013). Memasukkan efek dominan dalam model statistik dapat meningkatkan kualitas prediksi (Wellmann & Bennewitz, 2012). Pada bagian ini ditampilkan ekstensi persamaan (3.14) untuk memasukkan efek aditif dan dominasi bersama dalam NN. Selain itu, ditampilkan juga algoritma yang digunakan untuk fitting model ini menggunakan pendekatan EB, dengan strategi yang mirip dikembangkan oleh MacKay (1992, 1994).

Model NN yang mencakup efek aditif dan dominasi secara bersama diberikan oleh:

$$y_i = \mu + \sum_{s_a}^{s_d} w_k^a g \left( b_k^a + \sum_{j=1}^p x_{ij} \beta_j^{a[k]} \right) + \sum_{k=1}^{s_d} w_k^d g \left( b_k^d + \sum_{j=1}^p z_{ij} \beta_j^{d[k]} \right) + \varepsilon_i \quad (3.21)$$

## Bayesian regularized neural networks



**Gambar 3.6** Representasi Grafis dari NN yang diberikan dalam persamaan (3.21) (Gianola et al., 2013).

Dimana  $Z = \{z_{ij}\} \in \{0,1\}$  adalah matriks kejadian untuk efek dominasi dimensi  $n \times p$ ; jika marker digunakan,  $z_{ij} = 1$  jika SNP  $j$  untuk individu  $i$  heterozigot, dan  $z_{ij} = 0$  sebaliknya. Lebih lanjut,  $s_a$  dan  $s_d$  adalah jumlah neuron untuk komponen aditif dan dominan pada lapisan tersembunyi. Parameter yang tersisa adalah untuk persamaan (3.14). Gambar 3.6 menunjukkan representasi grafis dari NN dengan komponen dominan dan aditif. Perhatikan bahwa tidak diperbolehkan koneksi antara prediktor aditif dan neuron dominan, atau sebaliknya, sehingga jumlah parameter untuk estimasi berkurang secara signifikan. Maka dari itu, kemampuan perkiraan universal tidak hilang. Perhatikan juga bahwa rasio

$$h_a^2 = \frac{1}{2\alpha} / \frac{1}{2\alpha} + \frac{1}{2\delta} + \frac{1}{2\beta} \quad (3.22)$$

dan

$$h_d^2 = \frac{1}{2\delta} / \frac{1}{2\alpha} + \frac{1}{2\delta} + \frac{1}{2\beta} \quad (3.23)$$

dapat diartikan sebagai proporsi varian karena masing-masing komponen aditif dan dominan (lihat Gianola et al., 2011). Masukan dibagi dalam kelompok-kelompok; setiap kelompok *input* memiliki tingkat kebisingan sendiri dalam jaringan umpan-maju dengan beberapa lapisan neuron yang tidak terhubung.

#### 3.24.4 Estimasi Empiris Bayes

Untuk kesederhanaan, asumsikan bahwa parameter  $\mu$  dalam persamaan (3.21) telah dieliminasi dengan memusatkan pengamatan. Secara kondisional pada parameter jaringan  $t$ ,  $n$  fenotipe diasumsikan didistribusikan secara independen dengan kepadatan:

$$p(y|\varphi, \sigma_e^2) = \prod_{i=1}^n N(y_i|\varphi, \sigma_e^2) \quad (3.24)$$

di mana  $\varphi$  menunjukkan vektor dimensi  $t \times 1$  termasuk semua kekuatan koneksi, dan koefisien untuk efek aditif dan dominasi, serta bobot dan bias;  $\sigma_e^2$  adalah varian residual.

Misalkan  $\varphi = (\theta_a' \theta_d')'$ , di mana  $\theta_a$  menunjukkan vektor dimensi  $m \times 1$  dengan kekuatan untuk efek aditif, dan  $\theta_d$  menunjukkan vektor dimensi  $q \times 1$  dengan kekuatan untuk efek dominan. Perhatikan bahwa  $t = m + q$ . Dengan asumsi distribusi normal independen apriori untuk kekuatan koneksi:  $p(\varphi|\psi) = MN[0, bdiag(\sigma_a^2 I_{m \times m}, \sigma_d^2 I_{q \times q})]$ , dimana  $\psi = (\sigma_a^2, \sigma_d^2)'$  adalah varian dari kekuatan koneksi dan bobot untuk efek aditif dan dominan, dan  $bdiag(\cdot, \dots, \cdot)$  mewakili matriks diagonal blok (Gianola et al., 2013).

#### 3.24.5 Mode Posterior Bersyarat

Dengan asumsi bahwa parameter varians diketahui, kepadatan posterior kekuatan koneksi diberikan oleh



$$p(\varphi|y, \psi, \sigma_e^2) = \frac{p(y|\varphi, \sigma_e^2)p(\varphi|\psi)}{p(y|\psi, \sigma_e^2)} \quad (3.25)$$

Penyebut dalam persamaan (3.25) adalah kepadatan marginal data, dan diberikan oleh:

$$p(y|\psi, \sigma_e^2) = p(y|\varphi, \sigma_e^2)p(\varphi|\psi)d\varphi \quad (3.26)$$

$$= C \times \int \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^n e_i^2 - \frac{1}{2\sigma_a^2} \theta_a' \theta_a - \frac{1}{2\sigma_d^2} \theta_d' \theta_d \right\} d\varphi$$

dimana  $c = \left(\frac{1}{2\pi\sigma_e^2}\right)^{\frac{n}{2}} \left(\frac{1}{2\pi\sigma_a^2}\right)^{\frac{m}{2}} \left(\frac{1}{2\pi\sigma_d^2}\right)^{\frac{q}{2}}$  dan  $e_i = y_i - \hat{y}_i$  adalah kesalahan prediksi.

Seperti sebelumnya, EB terdiri dari 2 langkah: pertama mode posterior bersyarat dari kekuatan koneksi diperoleh dengan asumsi bahwa komponen varians diketahui, dan kemudian komponen varians ditaksir ulang pada langkah kedua. Dari persamaan (3.25) perhatikan bahwa:

$$p(\varphi|y, \psi, \sigma_e^2) \propto p(y|\varphi, \sigma_e^2)p(\varphi|\psi), \text{ sehingga} \quad (3.27)$$

$$\log p(\varphi|y, \psi, \sigma_e^2) \propto \log p(y|\varphi, \sigma_e^2) + \log p(\varphi|\psi)$$

$$= -\beta \sum_{i=1}^n e_i^2 - \alpha \theta_a' \theta_a - \delta \theta_d' \theta_d$$

Dimana  $\beta = \frac{1}{2\sigma_e^2}, \alpha = \frac{1}{2\sigma_a^2}, \delta = \frac{1}{2\sigma_d^2}$ .

Dengan asumsi  $\beta, \alpha,$  dan  $\delta$  diketahui, kita dapat memaksimalkan log-posterior ini untuk mendapatkan  $\varphi = \varphi^{map}$ . Perhatikan bahwa ini setara dengan meminimalkan:

$$Q(\varphi) = \beta \sum_{i=1}^n e_i^2 + \alpha \theta_a' \theta_a - \delta \theta_d' \theta_d \quad (3.28)$$

### 3.24.6 Tuning $\beta, \alpha, \delta$

Untuk menyetel parameter varians dapat memaksimalkan kemungkinan marginal dari data yang diberikan dalam persamaan (3.26). Menurut Tierney dan Kadane (1986) dalam jurnal Gianola et al. (2013) integral dalam persamaan itu tidak memiliki bentuk tertutup karena nonlinier, tetapi dapat diperkirakan dengan menggunakan metode Laplace yaitu:

$$\int \exp\{-nh(\varphi)\}d\varphi \approx \left(\frac{2\pi}{n}\right)^t |\Sigma|^{-1/2} \exp[-nh(\varphi)] \quad (3.29)$$

dimana  $t$  adalah jumlah elemen dalam vektor  $\varphi$ ,  $n$  adalah jumlah titik data, dan  $\Sigma$  adalah matriks Hessian, yaitu matriks turunan kedua  $h(\varphi)$  sehubungan dengan  $\varphi$ .

Menggunakan persamaan (3.29) dalam persamaan (3.26), integral  $I = \int \exp[-Q(\varphi)]d\varphi$  dapat diperkirakan sebagai berikut:

$$I \approx \left(\frac{2\pi}{n}\right)^t |\Sigma|^{-1/2} \exp[-Q(\varphi)] \quad (3.30)$$

Dimana  $\Sigma = \nabla^2 Q(\varphi) = \beta \nabla^2 \sum_{i=1}^n e_i^2 + \alpha \nabla^2 \theta_a' \theta_a + \delta \nabla^2 \theta_d' \theta_d$

Matriks Hessian dapat diperkirakan dengan menggunakan prosedur yang mirip dengan yang digunakan oleh (Foresee & Hagan, 1997). Perhatikan dulu itu

$$\beta \nabla^2 \sum_{i=1}^n e_i^2 \approx 2\beta J'J \quad (3.31)$$

dimana  $J$  adalah Jacobian untuk set kesalahan pelatihan (Gianola et al., 2013); kemudian  $\alpha \nabla^2 \theta_a' \theta_a = b \text{diag}(2\alpha I_{m \times m}, 0_{q+2 \times q+2})$ ,  $\delta \nabla^2 \theta_d' \theta_d = b \text{diag}(0_{m \times m}, 2\delta I_{q \times q}, 0_{2 \times 2})$  jadi  $\Sigma \approx 2\beta J'J + 2b \text{diag}(\alpha I_{m \times m}, \delta I_{q \times q})$ .

Mengganti persamaan (3.30) ke persamaan (3.26) dan mengambil logaritma alami, seseorang memiliki:

$$\begin{aligned} G(\beta, \alpha, \delta, \psi) &= \log p(y|\psi, \sigma_e^2) \\ &\approx c + \frac{n}{2} \log \beta + \frac{m}{2} \log \alpha + \frac{q}{2} \log \delta - \frac{1}{2} \log |\Sigma|_{\varphi=\varphi^{map}} \\ &\quad - Q(\varphi) |_{\varphi=\varphi^{map}} \end{aligned} \quad (3.32)$$

dimana  $c$  adalah konstanta.

Sekarang dapat diambil turunan parsial dari  $G(\cdot)$  sehubungan dengan masing-masing parameter dispersi,  $\beta$ ,  $\alpha$ ,  $\delta$  diatur ke 0, dan mendapatkan ekspresi untuk memperbarui parameter ini. Untungnya, masalah ini memiliki solusi yang persis sama dengan yang ditemukan oleh MacKay (1992, 1994) dalam konteks "konstanta regularisasi berganda" (yaitu, penugasan distribusi sebelum yang berbeda untuk bobot, bias, dan bias umum dalam NN). Diperoleh iterasi, dari kanan ke kiri:

$$\alpha_{new} = \gamma_a / 2\theta'_a \theta_a \text{ dengan } \gamma_a = m - 2a \text{Trace}_a(\Sigma^{-1}),$$

$$\delta_{new} = \gamma_d / 2\theta'_d \theta_d \text{ dengan } \gamma_d = q - 2\delta \text{Trace}_d(\Sigma^{-1}),$$

$$\beta_{new} = \frac{n-\gamma}{2 \sum_{i=1}^n e_i^2},$$

Jejak Hessian terbalik diambil atas bobot dan bias untuk aditif dan dominasi, masing-masing, dan  $\gamma = \gamma_a + \gamma_d$ .

MacKay (1992, 1994) juga memberikan perkiraan yang tidak memerlukan perhitungan para Hessian:

$$\alpha_{new} = \frac{m}{4\theta'_a \theta_a}, \delta_{new} = \frac{q}{4\theta'_d \theta_d} \text{ dan } \beta_{new} = \frac{n}{4 \sum_{i=1}^n e_i^2}$$

### 3.24.7 Algoritma

Algoritma untuk model fitting persamaan (3.21) dapat diringkas sebagai berikut:

1. Inisialisasi  $\beta$ ,  $\alpha$ ,  $\delta$  dan bobot menggunakan algoritma Nguyen dan Widrow (1990).
2. Ambil 1 langkah dari algoritma Levenberg-Marquardt untuk meminimalkan fungsi objektif  $Q(\varphi)$  yang diberikan dalam persamaan (3.28).
3. Perbarui  $\beta$ ,  $\alpha$ ,  $\delta$  dengan memaksimalkan persamaan (3.32) menggunakan algoritma Nelder dan Mead (1965) dalam jurnal Gianola et. Al (2013), atau menggunakan rumus pembaruan MacKay (1992, 1994), seperti yang ditunjukkan di atas.
4. Ulangi Langkah 2 dan 3 hingga konvergen.

Algoritma yang dijelaskan di atas adalah generalisasi dari algoritma yang disajikan dalam Foresee dan Hagan (1997). Peringatan yang ditunjukkan oleh Carlin dan Louis (2009) dalam jurnal yang ditulis oleh Gianola et al. (2013) adalah bahwa perkiraan Laplace mungkin tidak cukup akurat untuk sedang hingga sangat dimensional  $\theta$  (mis., Lebih besar dari 10).

Algoritma baru dapat diperluas untuk mengakomodasi situasi di mana entri hubungan aditif dan dominasi digunakan sebagai *input* dalam jaringan. persamaan (3.18), diperluas untuk memasukkan efek dominasi, menjadi:

$$y = u + d + \varepsilon \quad (3.33)$$

di mana  $d$  adalah vektor efek dominan,  $d \sim NM(0, \sigma_d^2 D)$ ,  $D$  adalah matriks hubungan dominan, dan  $\sigma_d^2$  adalah varian dominan.

Menggunakan representasi yang analog dengan yang digunakan oleh Gianola et al. (2011), model dapat ditulis sebagai  $y = u + d + \varepsilon = Lz\sigma_u + Wv\sigma_d + \varepsilon = Lu^* + Wd^* + \varepsilon$  dengan  $WW^T = D$  dan  $v \sim NM(0, I)$ , menghasilkan kasus khusus dari persamaan (3.21). Matriks hubungan dominasi genom  $D$  dapat dihitung menggunakan argumen yang sama dari Van Raden (2008) dalam jurnal Gianola et al. (2013) seperti:

$$D = \frac{ZZ'}{2 \sum_{j=1}^p p_j q_j (1 - 2p_j q_j)} \quad (3.34)$$

di mana  $q_j = 1 - p_j, j = 1, \dots, p$  dan  $Z = \{z_{ij}\} \in \{0, 1\}$  adalah matriks kejadian untuk efek dominasi dari penanda dimensi  $n \times p$ .

### 3.25 Package BRNN dalam Software Rstudio

Fungsi brnn cocok dengan jaringan saraf dua lapis seperti yang dijelaskan dalam MacKay (1992) dan Foresee and Hagan (1997). Ini menggunakan algoritma Nguyen dan Widrow (1990) untuk menetapkan bobot awal dan algoritma Gauss Newton untuk melakukan optimasi. Model ini diberikan oleh (Foresee & Hagan, 1997):

$$y_i = g(\{x\}_i) + e_i = \sum_{k=1}^s w_k g_k(b_k + \sum_{j=1}^p x_{ij} \beta_j^{[k]}) + e_i \quad (3.35)$$

dimana:

$e_i \sim N(0, \sigma_e^2), i = 1, \dots, n$

$s$  : jumlah neuron.

$w_k$  : bobot neuron k-th,  $k = 1, \dots, s$ .

$b_k$  : bias untuk neuron k-th,  $k = 1, \dots, s$ .

$\beta_j^{[k]}$  : bobot dari input j-th ke net,  $j = 1, \dots, p$ .

$g_k(\cdot)$  : fungsi aktivasi, dalam implementasi ini  $g_k(x) = \frac{\exp(2x)-1}{\exp(2x)+1}$ .

Perangkat lunak akan meminimalkan:

$$F = \beta E_d + \alpha E_w \quad (3.36)$$

dimana:

$y_i - \hat{y}_i$ , mis. jumlah kesalahan kuadrat.

$E_w$  : Jumlah kuadrat dari parameter jaringan (bobot dan bias).

$\beta$  :  $1 / (2\sigma_e^2)$ .

$\alpha$  :  $1 / (2\sigma_\theta^2)$ ,  $\sigma_\theta^2$  adalah parameter dispersi untuk bobot dan bias.

Objek dari kelas "brnn" atau "brnn.formula". Sebagian besar struktur internal, tetapi daftar yang berisi:

\$ theta : Daftar yang berisi bobot dan bias. Komponen pertama dari daftar berisi vektor dengan parameter estimasi untuk neuron k-th, yaitu  $(w_k, b_k, \beta_1^{\{k\}}, \dots, \beta_p^{\{k\}})$  :

\$ message : Tali yang menunjukkan kriteria berhenti untuk proses pelatihan.

\$ alpha : Parameter alfa.

\$ beta : Parameter beta.

\$ gamma : Jumlah parameter yang efektif.

\$ Ew : Jumlah kuadrat dari bias dan bobot.

\$ Ed : Jumlah kuadrat antara nilai yang diamati dan yang diprediksi.

### 3.26 Regresi

Istilah regresi pertama kali digunakan oleh *Francis Galton*. Interpretasi dari analisis regresi adalah analisis regresi berhubungan dengan studi mengenai ketergantungan dari sebuah variabel, yaitu variabel dependen, terhadap satu atau lebih variabel yang lain, yaitu variabel-variabel independen, dengan tujuan untuk menaksir dan/atau meramal rata-rata atau *mean* populasi variabel dependen dengan dasar nilai tertentu dari variabel penjelas (Hakim, 2004).

Secara umum regresi linear terdiri dari dua, yaitu regresi linear sederhana dengan satu buah variabel bebas dan satu buah variabel terikat; dan regresi linear berganda dengan beberapa variabel bebas dan satu buah variabel terikat. Analisis regresi linear merupakan metode statistik yang paling jamak dipergunakan dalam penelitian-penelitian sosial, terutama penelitian ekonomi.

Penelitian sebanyak  $n$  dengan variabel bebas ( $x$ ) sebanyak  $p$  maka model regresi dituliskan sebagai berikut (Walpole & Myers, 1995):

$$Y_i = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \dots + \beta_k X_{ik} + \varepsilon_i ; \quad i = 1, 2, \dots, n \quad (3.37)$$

dengan

$Y_i$  : nilai observasi variabel terikat ke- $i$

$X_{ik}$  : nilai observasi variabel bebas ke- $k$  pada pengamatan ke- $i$

$\beta_0$  : nilai intersep model regresi

$\beta_k$  : koefisien regresi variabel bebas ke- $k$

$\varepsilon_i$  : eror pada pengamatan ke- $i$

Pada pemodelan regresi terdapat asumsi-asumsi yang harus dipenuhi yaitu asumsi normalitas, asumsi homoskedastisitas, asumsi *non* autokorelasi, dan asumsi *non* multikolinieritas.

### 3.26.1 Uji Normalitas

Asumsi normal digunakan untuk mengetahui apakah residual berdistribusi normal. Jika asumsi kenormalan tidak terpenuhi, maka estimasi *Ordinary Least Square* (OLS) tidak dapat digunakan. Beberapa pengujian yang dapat dilakukan untuk asumsi distribusi normal adalah dengan normal P-P Plot, *Anderson Dalring*, *Kolmogorov-Smirnov*, *Jarque-Bera Test*, *Shapiro Wilk Test* dan *Skewnes-Kurtosis*. Hipotesis untuk uji *Kolmogorov-Smirnov* adalah:

$H_0$  : Residual berdistribusi normal

$H_1$  : Residual tidak berdistribusi normal

### 3.26.2 Uji Homoskedastisitas

Asumsi homoskedastisitas digunakan untuk melihat apakah varians eror homogen atau tidak. Pendeteksian penyimpangan asumsi homoskedastisitas ini dapat dilihat dari grafik plot nilai kuadrat residual. Jika nilai kuadrat residual membentuk pola yang sistematis maka dapat dikatakan terjadi heteroskedastisitas. Selain itu dapat juga dilakukan dengan pengujian *Breusch Pagan Godfrey*, *Harvey*, *Glejser*, ARCH, dan *White Test*. Hipotesis untuk uji *Breusch Pagan Godfrey* adalah sebagai berikut:

$H_0$  : Varians eror homogen (homoskedastisitas)

$H_1$  : Varians eror tidak homogen (heteroskedastisitas)

### 3.26.3 Uji Asumsi Non Autokorelasi

Autokorelasi adalah terjadinya korelasi antara satu variabel eror dengan variabel eror lainnya. Autokorelasi seringkali terjadi pada data *time series* dan dapat juga terjadi pada data *cross section* tetapi jarang (Widarjono, 2005). Asumsi autokorelasi digunakan untuk melihat apakah eror antar pengamatan berkorelasi atau tidak. Hal yang dilakukan untuk mendeteksi adanya autokorelasi dalam model regresi linier berganda adalah menggunakan metode *Durbin-Watson*. Hipotesis untuk uji *Durbin-Watson* adalah sebagai berikut:

$H_0$  :  $\rho = 0$  (Tidak terdapat autokorelasi)

$H_1$  :  $\rho \neq 0$  (Terdapat autokorelasi)

### 3.26.4 Uji Asumsi Non Multikolinearitas

Asumsi Multikolinearitas menunjukkan adanya hubungan linear yang kuat antara beberapa variabel bebas dalam suatu model regresi linear berganda. Model regresi yang baik memiliki variabel-variabel prediktor yang tidak berkorelasi. Kolinearitas terjadi karena terdapat korelasi yang cukup tinggi di antara variabel independen.

Masalah multikolinearitas dapat diketahui menggunakan nilai *Tolerance* (*TOL*) dan *Variance Inflation Factor* (*VIF*). Apabila nilai  $TOL < 0,1$  atau nilai  $VIF > 10$  maka dapat dikatakan terdapat masalah multikolinearitas. Rumus *VIF* didefinisikan sebagai berikut (Montgomery, Peck, & Vining, 1991) :

$$VIF = \frac{1}{1 - R_j^2} \quad (3.38)$$

dengan  $j = 1, 2, \dots, k$  dan  $k$  adalah banyaknya variabel independen, sedangkan  $R_j^2$  adalah koefisien determinasi yang dihasilkan dari regresi variabel independen  $X_j$  dengan variabel independen lain. Hipotesis untuk uji multikolinearitas adalah sebagai berikut:

$H_0$  : Tidak terdapat multikolinearitas

$H_1$  : Terdapat multikolinearitas



### 3.27 Parameter Evaluasi

Ada dua alasan utama untuk mengukur akurasi prediksi model *time series* (Vercellis, 2009). Pertama, pada tahap pengembangan dan identifikasi model, ukuran akurasi diperlukan untuk membandingkan model-model alternatif satu sama lain dan untuk menentukan nilai parameter yang muncul dalam ekspresi untuk fungsi prediksi  $F$ . Untuk mengidentifikasi model prediksi yang paling akurat, masing-masing model dianggap diterapkan pada data masa lalu, dan model dengan total eror minimum dipilih.

Kedua, setelah model prediksi telah dikembangkan dan digunakan untuk menghasilkan prediksi untuk masa mendatang, perlu untuk secara berkala menilai keakuratan, untuk mendeteksi kelainan dan kekurangan dalam model yang mungkin timbul di lain waktu. Evaluasi keakuratan prediksi pada tahap ini membuat mungkin untuk menentukan apakah model masih akurat atau memerlukan suatu revisi. Untuk mengevaluasi akurasi dan peramalan kinerja model berbeda, penelitian ini mengadopsi dua indeks evaluasi: *Mean Square Error* (MSE) atau *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Formula untuk menghitung indeks ini diberikan di bawah ini (Budiman, 2016):

- *Mean Square Error* (MSE)

MSE adalah penjumlahan kuadrat eror atau selisih antara nilai sebenarnya (aktual) dan nilai prediksi, kemudian membagi jumlah tersebut dengan banyaknya waktu data peramalan, atau dapat dirumuskan sebagai berikut:

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \quad (3.39)$$

- *Root Mean Square Error* (RMSE)

RMSE adalah penjumlahan kuadrat eror atau selisih antara nilai sebenarnya (aktual) dan nilai prediksi, kemudian membagi jumlah tersebut dengan banyaknya waktu data peramalan dan kemudian menarik akarnya, atau dapat dirumuskan sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (3.40)$$



- *Mean Absolute Percentage Error (MAPE)*

MAPE adalah nilai absolute dari persentase eror data terhadap *mean*, atau dapat dirumuskan sebagai berikut:

$$MAPE = \frac{\sum_{i=1}^n \frac{|\hat{Y}_i - Y_i|}{Y_i} \times 100}{n} \quad (3.41)$$

