



BAB II LANDASAN TEORI

© Hak cipta milik IBI KIK (Institut Bisnis dan Informatika Kwik Kian Gie)

A. Sistem Informasi

Menurut Kenneth C. Laudon dan Jane P. Laudon (2017:44) sistem informasi dapat didefinisikan secara teknis sebagai satu set komponen yang mengumpulkan (atau mengambil), memproses, menyimpan, dan mendistribusikan informasi untuk mendukung pengambilan keputusan dan pengendalian dalam suatu organisasi.

Menurut Ralph M. Stair dan George W. Reynolds (2017:6) Sistem Informasi (SI) adalah seperangkat komponen yang saling terkait yang mengumpulkan, memproses, menyimpan, dan menyebarluaskan data dan informasi.

Menurut Joseph Valacich dan Christoph Schneider (2017:16) Sistem Informasi (SI) adalah kombinasi orang dan teknologi informasi yang membuat, mengumpulkan, memproses, menyimpan, dan mendistribusikan data yang berguna.

B. Data

Menurut Kenneth C. Laudon dan Jane P. Laudon (2017:44) data adalah aliran fakta mental yang merepresentasikan kejadian yang terjadi pada suatu organisasi atau lingkungan fisik sebelum diorganisir dan disusun menjadi bentuk yang dapat dimengerti dan digunakan oleh seseorang.

Menurut Ralph M. Stair dan George W. Reynolds (2017:4) data terdiri dari fakta mental, seperti nomor karyawan, jumlah jam kerja dalam seminggu, jumlah bagian persediaan, atau jumlah unit yang diproduksi pada garis produksi.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Menurut Patricia Wallace (2015:9) data adalah fakta individual atau bagian dari sebuah informasi.

Menurut Joseph Valacich dan Christopher Schenider (2017:18) data adalah simbol mental, seperti karakter dan angka.

C. Informasi

Ralph M. Stair dan George W. Reynolds (2017:4) mengartikan bahwa Informasi adalah kumpulan data yang diatur dan diproses sehingga memiliki nilai tambahan di luar nilai fakta individu.

Menurut Kenneth C. Laudon dan Jane P. Laudon (2017:44) informasi adalah data yang telah dibentuk menjadi bentuk yang berarti dan berguna untuk manusia.

Menurut Joseph Valacich dan Christoph Schneider (2017:19) informasi adalah data dapat diformat, diorganisasikan, atau diproses agar bermanfaat.

D. Database

Patricia Wallace (2015:102) mengartikan *database* sebagai sebuah kumpulan informasi terintegrasi yang secara logis terkait dan disimpan sedemikian rupa untuk meminimalkan duplikasi dan memfasilitasi pengambilan secara cepat.

Menurut Joseph Valacich dan Christoph Schneider (2017:96) *database* merupakan kumpulan data terkait yang diatur agar mampu memfasilitasi pencarian data,

Ralph M. Stair dan George W. Reynolds (2017:4) mengartikan *database* sebagai kumpulan fakta dan informasi terorganisir yang biasanya terdiri dari dua atau lebih file data terkait

Hak Cipta Milik IBIKKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Menurut Kenneth C. Laudon dan Jane P. Laudon (2017:245) database adalah kumpulan data yang diatur untuk melayani banyak aplikasi secara efisien dengan memusatkan data dan mengendalikan data yang berlebihan.

E. Database Management Systems (DBMS)

Menurut Kenneth C. Laudon dan Jane P. Laudon (2017:245) Database Management Systems (DBMS) adalah kumpulan data yang diatur untuk melayani banyak aplikasi secara efisien dengan memusatkan data dan mengendalikan data yang berlebihan.

Menurut Joseph Valacich dan Christoph Schneider (2017:231) DBMS adalah aplikasi perangkat lunak yang digunakan untuk membuat, menyimpan, mengatur, dan mengambil data dari database tunggal atau beberapa basis data.

F. Data Warehouse

Menurut Kenneth C. Laudon dan Jane P. Laudon (2017:255) data warehouse adalah database yang menyimpan data historis saat ini yang berpotensi untuk pengambilan keputusan di perusahaan.

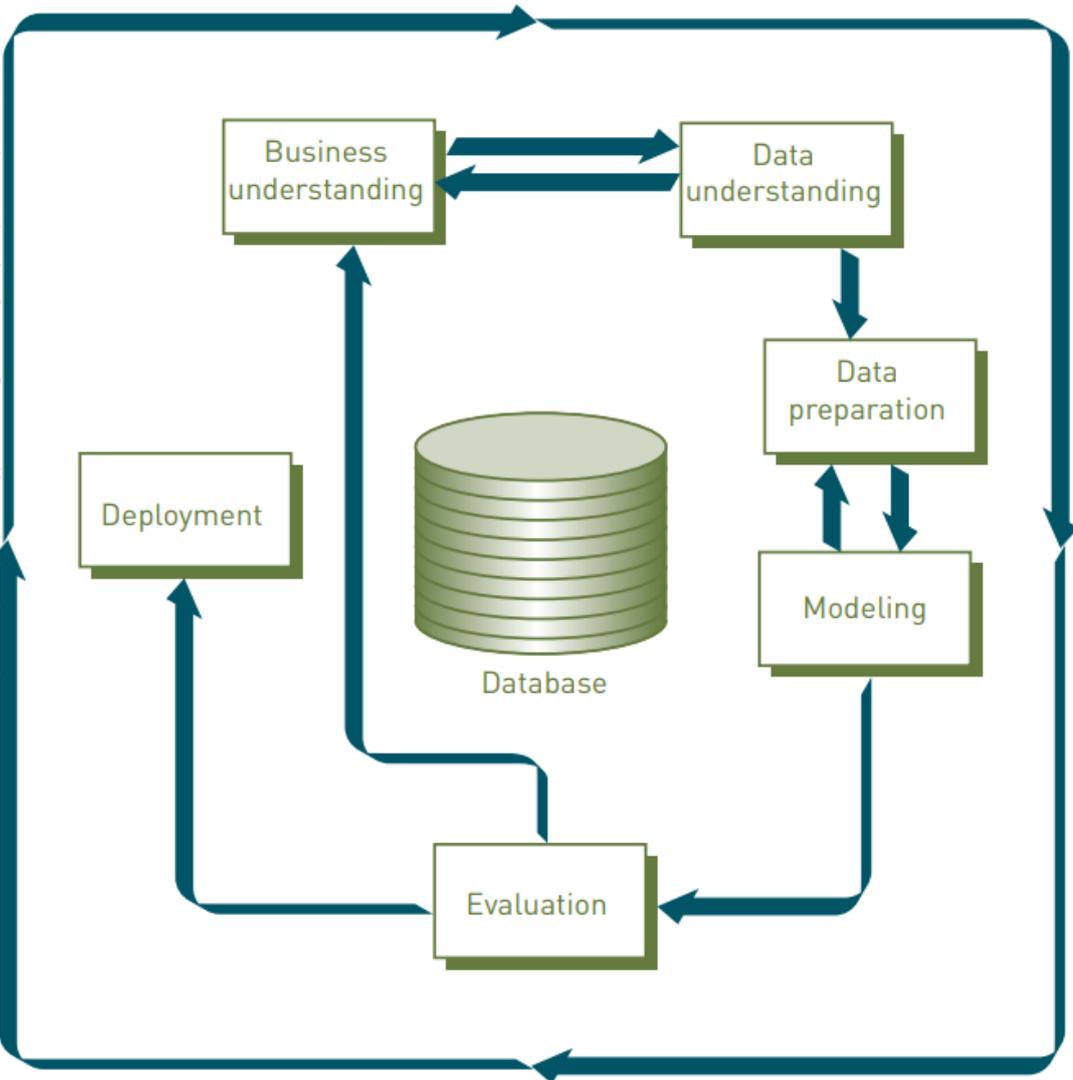
Menurut Joseph Valacich dan Christoph Schneider (2017:238) data warehouse adalah gudang data yang mengintegrasikan beberapa basis data besar dan sumber data lainnya ke dalam satu gudang. Gudang tersebut berisi data historis dan data (hampir) terkini untuk analisis dan pelaporan yang cocok untuk digunakan bila ada permintaan data, analisis, atau pemrosesan langsung.

Patricia Wallace (2015:114) mengartikan data warehouse sebagai pusat penyimpanan data yang berisi informasi yang diambil dari berbagai sumber yang dapat digunakan untuk analisis, pengumpulan intelijen, dan perencanaan strategis.

Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

metodologi yang kuat dan terbukti dengan baik, dan meskipun pertama kali disusun pada tahun 1999 metodologi ini tetap menjadi yang paling banyak digunakan untuk proyek-proyek penambangan data. Tahapan ini bisa dilihat pada gambar dibawah:

Gambar 2.1
Tahapan *Data Mining* CRISP-DM



Sumber : Ralph M. Stair dan George W. Reynolds (2017:394)

Setiap tahapan dalam metode ini memiliki tujuan dan proses yang berbeda yaitu:

1. *Business Understanding* (pemahaman bisnis): Memperjelas tujuan bisnis untuk proyek penambangan data, konversi sasaran menjadi masalah analisis prediksi, dan merancang proyek yang tepat untuk mencapai tujuan tersebut.





2. *Data Understanding* (pemahaman atau pengertian data): Mengumpulkan data yang akan digunakan (yang mungkin melibatkan banyak sumber), proses pembiasaan data, dan mengidentifikasi masalah kualitas data (kurangnya data, data yang hilang, penyesuaian kebutuhan data, dll.) serta kebutuhan data lainnya yang perlu ditangani.
3. *Data Preparation* (persiapan data): Pilih subkumpulan data yang akan digunakan, bersihkan data untuk mengatasi masalah kualitas, dan ubah data menjadi bentuk yang cocok untuk analisis.
4. *Modeling* (pemodelan): Terapkan teknik pemodelan yang dipilih.
5. *Evaluation* (evaluasi): Menilai apakah model mencapai sasaran atau tujuan yang diinginkan.
6. *Deployment* (penyebaran): erapkan model ke dalam proses pengambilan keputusan.

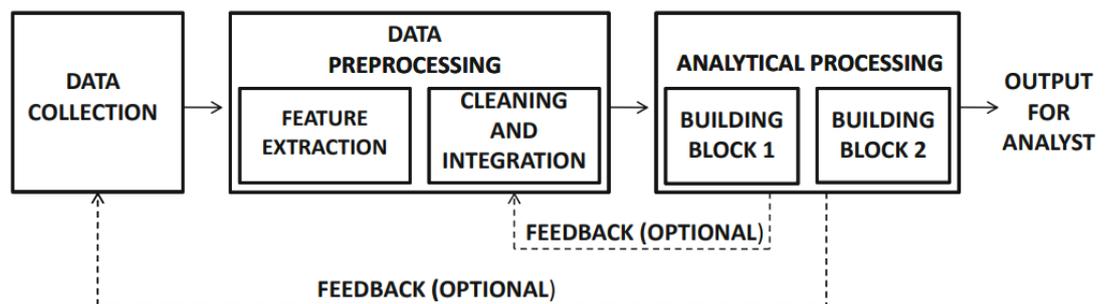
C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

Menurut Charu C Aggarwal (2015:3) tahapan dalam data mining dibagi dalam 3 proses besar yaitu :

Gambar 2.2
Tahapan Data Mining Pipeline Model



Sumber : Charu C Aggarwal (2015:4)



1. *Data collection* (Pengumpulan data) : Tahap dimana data yang dibutuhkan untuk penelitian dikumpulkan sesuai dengan kebutuhan. Pengumpulan data mungkin memerlukan penggunaan hardware khusus seperti jaringan sensor, tenaga kerja manual seperti koleksi survei pengguna, atau perangkat lunak seperti mesin dokumen perangkak web (*web crawling*) untuk mengumpulkan dokumen. Sementara tahap ini sangat-aplikasi spesifik dan sering di luar bidang *analisis data mining*, hingga menyebabkan tahap ini sangat penting karena pilihan yang baik pada tahap ini secara signifikan dapat mempengaruhi proses *data mining*. Setelah tahap pengumpulan, data sering disimpan dalam *database*, atau, lebih umum, sebuah gudang data untuk diproses.

2. *Feature extraction and data cleaning* (Ekstraksi fitur dan pembersihan data) : Tahap dimana data di ekstraksi, seleksi, dan dibersihkan untuk digunakan dalam pemodelan. Ketika data dikumpulkan, mereka sering tidak dalam bentuk yang cocok untuk diproses. Sebagai contoh, data dapat dikodekan dalam log kompleks atau dokumen-bentuk bebas. Dalam banyak kasus, jenis data yang berbeda dapat sewenang-wenang dicampur bersama dalam dokumen dengan bentuk bebas. Untuk membuat data yang sesuai untuk pengolahan, penting untuk mengubah mereka menjadi format yang ramah untuk algoritma *data mining*, seperti multidimensi, *time series*, atau format semi terstruktur.

Format multidimensi adalah salah satu yang paling umum, di mana berbagai bidang data sesuai dengan sifat yang diukur berbeda yang disebut sebagai

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



fitur, atribut, atau dimensi. Hal ini penting untuk mengekstrak fitur yang relevan untuk proses penambangan. Tahap ekstraksi fitur sering dilakukan secara paralel dengan pembersihan data, di mana dalam proses tersebut akan menghilangkan bagian yang salah dari data yang baik. Dalam banyak kasus, data dapat diekstraksi dari berbagai sumber dan perlu diintegrasikan ke dalam format terpadu untuk diproses. Hasil akhir dari prosedur ini adalah kumpulan data terstruktur dengan baik, yang dapat secara efektif digunakan oleh program komputer. Setelah tahap ekstraksi fitur, data dapat lagi disimpan dalam *database* untuk diproses.

3. *Analytical processing and algorithm* : Proses analisa dan aplikasi algoritma pada data yang telah dibersihkan. Bagian akhir dari proses *data mining* adalah untuk merancang metode analisis yang efektif dari data yang telah diproses.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

I. Teknik-Teknik *Data Mining*

Menurut David M Kroenke (2014:241) dalam *data mining* teknik yang dimiliki terbagi menjadi 2 macam yaitu :

1. *Unsupervised Data Mining* : Analisis tidak membuat model atau hipotesis sebelum menjalankan analisis. Sebaliknya, mereka menerapkan teknik *data mining* untuk data dan mengamati hasil. Dengan metode ini, analisis membuat hipotesis setelah analisis menjelaskan pola yang ditemukan. Salah satu teknik tanpa pengawasan umum adalah analisis kluster (*cluster analysis*). Dengan itu, teknik statistik mengidentifikasi kelompok entitas yang memiliki karakteristik serupa. Yang umum digunakan untuk analisis kluster adalah



untuk menemukan kelompok pelanggan yang sama dari pesanan pelanggan dan data demografis.

2. *Supervised Data Mining* : Penambang data yang mengembangkan model sebelum analisis dan menerapkan teknik statistik data untuk memperkirakan parameter dari model. Misalnya, ahli pemasaran di sebuah perusahaan komunikasi percaya bahwa penggunaan ponsel pada akhir pekan ditentukan oleh usia pelanggan dan jumlah bulan saat pelanggan telah memiliki akun ponsel. Seorang analis *data mining* kemudian akan menjalankan analisis yang memperkirakan dampak pelanggan dan rekening usia. Salah satu analisis tersebut, yang mengukur dampak dari set variabel pada variabel lain, disebut analisis regresi (*regression analysis*).

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

J. Metode Dalam Data Mining

Menurut Jiawei Han, Micheline Kamber, dan Jian Pei (2012:15) ada beberapa jenis fungsionalitas database yang dapat yang dapat dilakukan dalam *data mining* antara lain adalah:

1. *Class/Concept Description: Characterization and Discrimination* (Deskripsi kelas/konsep : karakterisasi dan diskriminasi) : Entri data yang dapat dikaitkan dengan kelas atau konsep. Fungsi ini terdiri dari :
 - a. *Data Characterization* (Karakterisasi Data) : adalah ringkasan dari karakteristik umum atau fitur dari kelas target data.
 - b. *Data Discrimination* (Diskriminasi Data) : adalah perbandingan fitur umum dari objek data kelas sasaran



terhadap fitur umum objek dari satu atau beberapa kelas yang kontras.

2. *Mining Frequent Patterns Associations, and Correlations* (Penggalian pola yang sering muncul : asosiasi dan korelasi) : Meneliti pola yang sering terjadi di data. Fungsi terdiri dari :
 - a. *Associations* (Asosiasi) : Pola dimana suatu variabel memiliki *confidence* (tingkat keyakinan) dengan variabel lain dan *support* (tingkat pendukung) dimana variabel memiliki pola yang sama.
 - b. *Correlations* (Korelasi) : Tingkat hubungan yang dimiliki oleh suatu variabel dengan variabel lain.
3. *Classifications and Regression for Predictive Analysis* (Klasifikasi dan regresi untuk analisis prediksi) :
 - a. *Classification* (Klasifikasi) : adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep.
 - b. *Regression* (Regresi) : adalah proses untuk mengestimasi hubungan antara variabel.

I. Indeks Pembangunan Manusia

Menurut *Wikipedia* Indeks Pembangunan Manusia (IPM) / *Human Development Index* (HDI) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan

dan standar hidup untuk semua negara seluruh dunia. IPM digunakan untuk mengklasifikasikan apakah sebuah negara adalah negara maju, negara berkembang atau negara terbelakang dan juga untuk mengukur pengaruh dari kebijaksanaan ekonomi terhadap kualitas hidup

Menurut Badan Pusat Statistik (BPS) IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya.

IPM dibentuk oleh 3 (tiga) dimensi dasar:

1. Umur panjang dan hidup sehat.
2. Pengetahuan.
3. Standar hidup layak.

Dalam penerapannya IPM memiliki beberapa manfaat seperti:

1. IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk).
2. IPM dapat menentukan peringkat atau level pembangunan suatu wilayah/negara.
3. Bagi Indonesia, IPM merupakan data strategis karena selain sebagai ukuran kinerja Pemerintah, IPM juga digunakan sebagai salah satu alokator penentuan Dana Alokasi Umum (DAU).

Untuk menghitung IPM, kita bisa menggunakan rumus berikut:

$$x - index = \frac{\min(x)}{\max(x)}$$





di mana min (x) dan max (x) adalah variabel angka maksimum dan minimum (x) yang dapat diperoleh. IPM menggambarkan tiga indikator umum berikut:

$$\text{Indeks Harapan Hidup} = \frac{LE - 25}{85 - 25}$$

$$\text{Indeks Pendidikan} = \frac{2}{3} \times ALI + \frac{1}{3} \times GER$$

$$\text{Angka Melek Huruf Dewasa (ALI)} = \frac{ALR - 0}{100 - 0}$$

$$\text{Gross Enrollment Ratio (GER)} = \frac{CGER - 0}{100 - 0}$$

$$\text{Indeks PDB} = \frac{\log(GDPpc) - \log(100)}{\log(40000) - \log(100)}$$

Keterangan:

LE: Angka harapan hidup

ALR: Angka melek huruf

CGER: Combined gross enrollment ratio

GDPpc: PDB perkapita berdasarkan PPP dalam USD

Namun, pada tahun 2010 terdapat perubahan pada metodologi penghitungan IPM.

Alasan yang dijadikan dasar perubahan metodologi penghitungan IPM yaitu

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



1. Beberapa indikator sudah tidak tepat untuk digunakan dalam penghitungan IPM. Angka melek huruf sudah tidak relevan dalam mengukur pendidikan secara utuh karena tidak dapat menggambarkan kualitas pendidikan. Selain itu, karena angka melek huruf di sebagian besar daerah sudah tinggi, sehingga tidak dapat membedakan tingkat pendidikan antardaerah dengan baik.
2. PDB per kapita tidak dapat menggambarkan pendapatan masyarakat pada suatu wilayah.
3. Penggunaan rumus rata-rata aritmatik dalam penghitungan IPM menggambarkan bahwa capaian yang rendah di suatu dimensi dapat ditutupi oleh capaian tinggi dari dimensi lain.

© Hak cipta milik IBI RKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

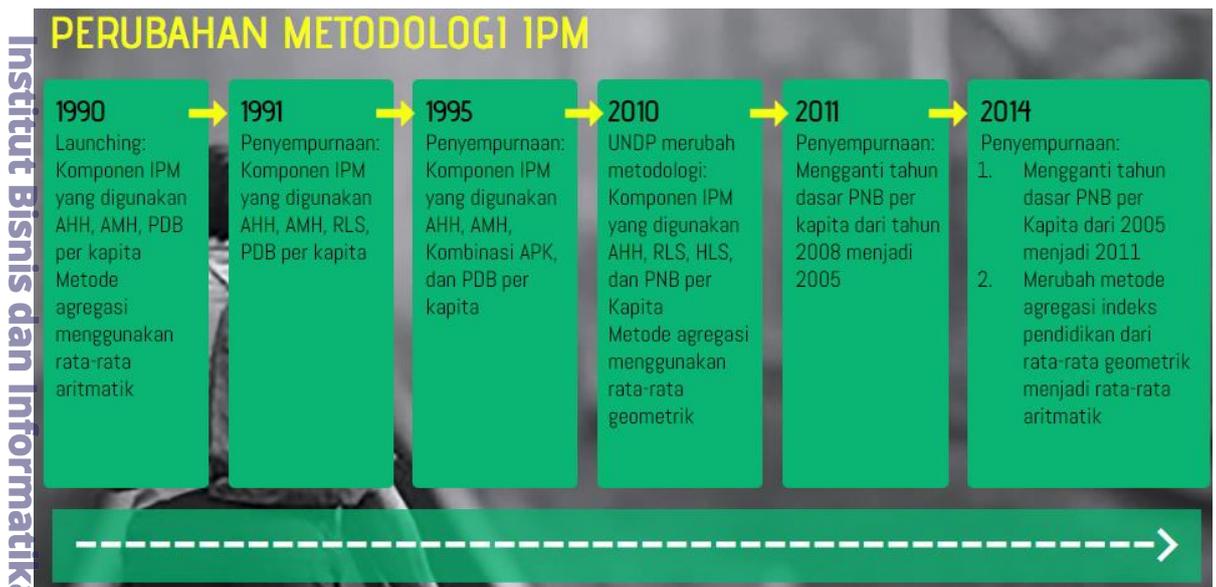
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IBI RKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBI RKG.

Pada tahun 2010 terdapat perubahan pada metodologi penghitungan IPM. Alasan yang dijadikan dasar perubahan metodologi penghitungan IPM yaitu

Sejarah perubahan metodologi IPM dapat terlihat pada gambar berikut:

Gambar 2.3
Sejarah Perubahan Metodologi IPM



Sumber : <http://ipm.bps.go.id/page/ipm>



Hak Cipta Dilindungi Undang-Undang

Dalam penerapan metodologi terbaru ini ada beberapa hal yang berubah yaitu:

1. Indikator

- a. Angka melek huruf pada metode lama diganti dengan Angka Harapan Lama Sekolah.
- b. Produk Domestik Bruto (PDB) per kapita diganti dengan Produk Nasional Bruto (PNB) per kapita.

2. Metode Penghitungan

Metode agregasi diubah dari rata-rata aritmatik menjadi rata-rata geometrik.

Dengan menggunakan metode baru ini ada beberapa keuntungan yang bisa diperoleh seperti:

1. Menggunakan indikator yang lebih tepat dan dapat membedakan dengan baik (diskriminatif).
 - a. Dengan memasukkan rata-rata lama sekolah dan angka harapan lama sekolah, bisa didapatkan gambaran yang lebih relevan dalam pendidikan dan perubahan yang terjadi.
 - b. PNB menggantikan PDB karena lebih menggambarkan pendapatan masyarakat pada suatu wilayah.
2. Penggunaan rata-rata geometrik dalam menyusun IPM dapat diartikan bahwa capaian satu dimensi tidak dapat ditutupi oleh capaian di dimensi lain. Artinya, untuk mewujudkan pembangunan manusia yang baik, ketiga dimensi harus memperoleh perhatian yang sama besar karena sama pentingnya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Berikut ini adalah gambar tabel perbandingan metodologi penghitungan IPM yang lama dengan yang baru.

Gambar 2.4
Perbandingan Metodologi IPM

Dimensi	Metode Lama		Metode Baru	
	UNDP	BPS	UNDP	BPS
Kesehatan	Angka Harapan Hidup saat Lahir (AHH)	Angka Harapan Hidup saat Lahir (AHH)	Angka Harapan Hidup saat Lahir (AHH)	Angka Harapan Hidup saat Lahir (AHH)
Pengetahuan	Angka Melek Huruf (AMH)	Angka Melek Huruf (AMH)	Harapan Lama Sekolah (HLS)	Harapan Lama Sekolah (HLS)
	Kombinasi Angka Partisipasi Kasar (APK)	Rata-rata Lama Sekolah (RLS)	Rata-rata Lama Sekolah (RLS)	Rata-rata Lama Sekolah (RLS)
Standar Hidup Layak	PDB per kapita (PPP US\$)	Pengeluaran per kapita Disesuaikan (Rp)	PNB per kapita (PPP US\$)	Pengeluaran per kapita Disesuaikan (Rp)
Agregasi	Rata-rata Aritmatik $IPM = \frac{1}{3}(I_{\text{kesehatan}} + I_{\text{pendidikan}} + I_{\text{pengeluaran}}) \times 100$		Rata-rata Geometrik $IPM = \sqrt[3]{I_{\text{kesehatan}} \times I_{\text{pendidikan}} \times I_{\text{pengeluaran}}} \times 100$	

Sumber : <http://ipm.bps.go.id/page/ipm>

Dalam menghitung IPM, diperlukan nilai minimum dan maksimum untuk masing-masing indikator. Berikut tabel yang menyajikan nilai-nilai tersebut.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Gambar 2.5
Tabel Indikator Variabel IPM

Indikator	Satuan	Minimum		Maksimum	
		UNDP	BPS	UNDP	BPS
Angka Harapan Hidup Saat Lahir	Tahun	20	20	85	85
Angka Harapan Lama Sekolah	Tahun	0	0	18	18
Rata-rata Lama Sekolah	Tahun	0	0	15	15
Pengeluaran per Kapita Disesuaikan		100 (PPP US\$)	1.007.436 * (Rp)	107.721 (PPP US\$)	26.572.352 ** (Rp)

Sumber : <http://ipm.bps.go.id/page/ipm>

Keterangan:

* **Daya beli minimum merupakan garis kemiskinan terendah kabupaten tahun 2010 (data empiris) yaitu di Tolikara-Papua**

** **Daya beli maksimum merupakan nilai tertinggi kabupaten yang diproyeksikan hingga 2025 (akhir RPJPN) yaitu perkiraan pengeluaran per kapita Jakarta Selatan tahun 2025**

Maka sesuai dengan tabel diatas, berikut ini adalah jenis variabel yang terdapat pada metode penghitungan IPM yang baru.

1. Angka Harapan Hidup Saat Lahir - AHH (Life Expectancy - e0):

Angka Harapan Hidup saat Lahir didefinisikan sebagai rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir. AHH mencerminkan derajat kesehatan suatu masyarakat. AHH dihitung dari hasil sensus dan survei kependudukan.

2. Rata-rata Lama Sekolah - RLS (Mean Years of Schooling - MYS):

Rata-rata Lama Sekolah didefinisikan sebagai jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal. Diasumsikan bahwa dalam kondisi normal rata-rata lama sekolah suatu wilayah tidak akan turun. Cakupan penduduk yang dihitung dalam penghitungan rata-rata lama sekolah adalah penduduk berusia 25 tahun ke atas.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

© Hak Cipta Dilindungi Undang-Undang
 IBIKKG (Institut Bisnis dan Informatika Kwik Kian Gie)



3. Angka Harapan Lama Sekolah - HLS (Expected Years of Schooling - EYS)

Angka Harapan Lama Sekolah didefinisikan lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang.

Diasumsikan bahwa peluang anak tersebut akan tetap bersekolah pada umur-umur berikutnya sama dengan peluang penduduk yang bersekolah per jumlah penduduk untuk umur yang sama saat ini.

Angka Harapan Lama Sekolah dihitung untuk penduduk berusia 7 tahun ke atas.

HLS dapat digunakan untuk mengetahui kondisi pembangunan sistem pendidikan di berbagai jenjang yang ditunjukkan dalam bentuk lamanya pendidikan (dalam tahun) yang diharapkan dapat dicapai oleh setiap anak.

4. Pengeluaran per Kapita Disesuaikan:

Pengeluaran per kapita yang disesuaikan ditentukan dari nilai pengeluaran per kapita dan paritas daya beli (Purchasing Power Parity-PPP). Rata-rata pengeluaran per kapita setahun diperoleh dari Susenas, dihitung dari level provinsi hingga level kab/kota. Rata-rata pengeluaran per kapita dibuat konstan/riil dengan tahun dasar

2012=100. Perhitungan paritas daya beli pada metode baru menggunakan 96

komoditas dimana 66 komoditas merupakan makanan dan sisanya merupakan

komoditas nonmakanan. Metode penghitungan paritas daya beli menggunakan

Metode Rao.

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IBI RKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

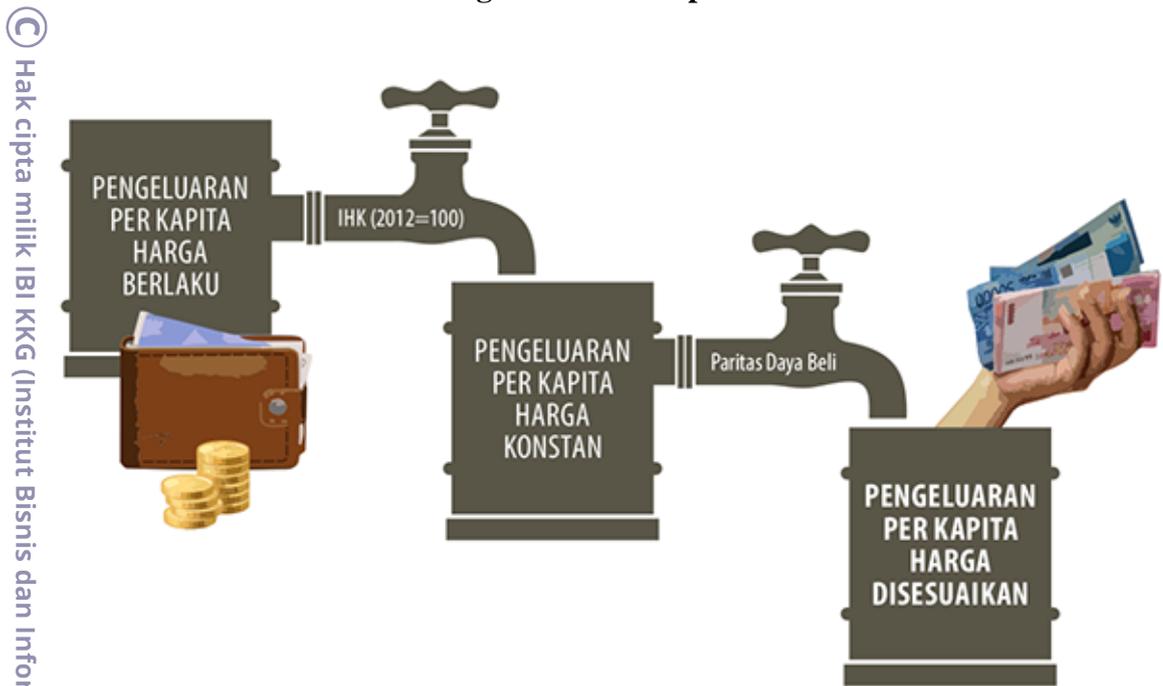
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

Gambar 2.6
Ilustrasi Pengeluaran Perkapita



Sumber : <http://ipm.bps.go.id/page/ipm>

Sedangkan untuk perhitungan paritas daya beli (PPP) Digunakan rumus berikut:

$$PPP_J = \prod_{i=1}^m \left(\frac{P_{ij}}{P_{ik}} \right)^{1/m}$$

Keterangan:

P_{ij} = Harga komoditas i di kab/kota j.

P_{ik} = Harga komoditas i di Jakarta Selatan.

m = jumlah komoditas.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Bila data variabel IPM yang dibutuhkan sudah didapat, maka selanjutnya penghitungan IPM dapat dilakukan dengan rumus berikut dimana Setiap komponen IPM distandardisasi dengan nilai minimum dan maksimum sebelum digunakan untuk menghitung IPM.

Dimensi Kesehatan:

$$I_{Kesehatan} = \frac{AHH - AHH_{min}}{AHH_{maks} - AHH_{min}}$$

Keterangan:

AHH = Angka Harapan Hidup

AHH_{maks} = Angka Harapan Hidup Tertinggi

AHH_{min} = Angka Harapan Hidup Terendah

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Dimensi Pendidikan:

$$I_{HLS} = \frac{HLS - HLS_{min}}{HLS_{maks} - HLS_{min}}$$

Keterangan:

HLS = Angka Harapan Lama Sekolah

HLS_{maks} = Angka Harapan Lama Sekolah Tertinggi

HLS_{min} = Angka Harapan Lama Sekolah Terendah

$$I_{RLS} = \frac{RLS - RLS_{min}}{RLS_{maks} - RLS_{min}}$$

Keterangan:

RLS = Rata-Rata Lama Sekolah

RLS_{maks} = Rata-Rata Lama Sekolah Tertinggi

RLS_{min} = Rata-Rata Lama Sekolah Terendah

$$I_{pendidikan} = \frac{I_{HLS} - I_{RLS}}{2}$$

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



$$I_{\text{pengeluaran}} = \frac{\ln(\text{pengeluaran}) - \ln(\text{pengeluaran}_{\min})}{\ln(\text{pengeluaran}_{\max}) - \ln(\text{pengeluaran}_{\min})}$$

Keterangan:

$\ln(\text{pengeluaran})$ = Pengeluaran per Kapita Disesuaikan

$\ln(\text{pengeluaran}_{\max})$ = Pengeluaran per Kapita Disesuaikan Tertinggi

$\ln(\text{pengeluaran}_{\min})$ = Pengeluaran per Kapita Disesuaikan Terendah

Bila perhitungan ketiga dimensi tersebut sudah terpenuhi, maka IPM bisa dihitung dengan menggunakan rumus berikut:

$$IPM = \sqrt[3]{I_{\text{kesehatan}} \times I_{\text{pendidikan}} \times I_{\text{pengeluaran}}}$$

L. Neural Network

Charu C Aggarwal (2015:326) mendefinisikan *neural network* sebagai model simulasi sistem saraf manusia. Sistem saraf manusia terdiri dari sel-sel, yang disebut sebagai *neuron*. *Neuron* biologis terhubung satu sama lain pada titik-titik kontak, yang disebut sebagai sinapsis. Pembelajaran dilakukan dalam organisme hidup dengan mengubah kekuatan koneksi sinaptik antar neuron. Biasanya, kekuatan koneksi ini berubah sebagai respons terhadap rangsangan eksternal. Jaringan syaraf tiruan dapat dianggap sebagai simulasi proses biologis ini.



Seperti dalam kasus jaringan biologis, *node* individu dalam jaringan saraf tiruan disebut sebagai *neuron*. *Neuron-neuron* ini adalah unit-unit komputasi yang menerima masukan dari beberapa *neuron* lain, membuat perhitungan pada input-input ini, dan memasukkannya ke dalam neuron-neuron lain. Fungsi perhitungan pada *neuron* ditentukan oleh bobot pada koneksi input ke *neuron* itu. Bobot ini dapat dilihat sebagai analog dengan kekuatan koneksi sinaptik. Dengan mengubah bobot ini secara tepat, fungsi perhitungan dapat dipelajari yang secara analog mirip dengan pembelajaran kekuatan sinaptik dalam jaringan saraf biologis. "Stimulus eksternal" dalam jaringan saraf tiruan untuk mempelajari bobot ini disediakan oleh data pelatihan. Identy adalah secara bertahap memodifikasi bobot kapan pun prediksi yang salah dibuat oleh kumpulan bobot saat ini.

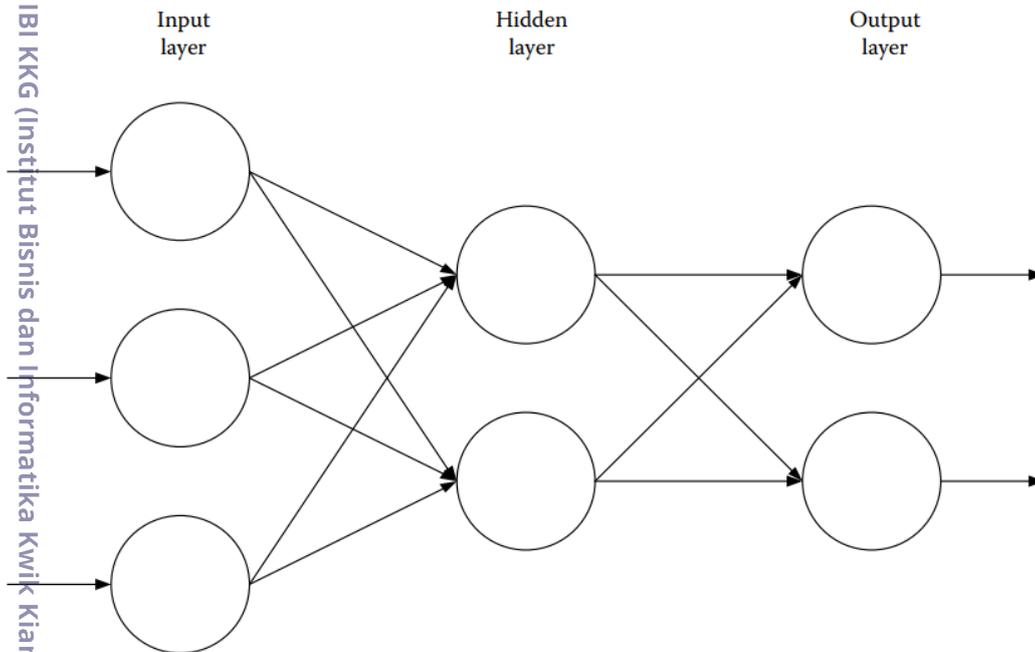
Menurut Richard J. Roiger (2016:5) *neural network* adalah seperangkat nodus yang saling berhubungan yang dirancang untuk meniru fungsi otak manusia. Karena otak manusia mengandung miliaran *neuron* dan jaringan saraf yang khas yang memiliki kurang dari 100 node, perbandingannya agak dangkal. Namun, jaringan saraf telah berhasil diterapkan untuk masalah di beberapa disiplin dan karena alasan ini cukup populer di komunitas penambangan data.

Neural Network hadir dalam berbagai bentuk dan dapat dikonstruksi untuk pembelajaran yang diawasi serta pengelompokan tanpa pengawasan. Dalam semua kasus, nilai yang dimasukkan ke jaringan saraf harus numerik. Jaringan umpan maju adalah model belajar yang diawasi yang populer.

Gambar 2.7 menunjukkan jaringan syaraf umpan-maju yang terhubung penuh yang terdiri dari tiga lapisan. Dengan jaringan umpan-maju, nilai-nilai atribut input untuk contoh individu masuk pada lapisan input dan lulus langsung melalui lapisan output dari struktur jaringan. Lapisan output mungkin berisi satu atau beberapa node. Lapisan output dari

jaringan yang ditunjukkan pada Gambar 2.7 berisi dua node. Oleh karena itu, *output* dari jaringan saraf akan menjadi pasangan nilai yang terurut.

Gambar 2.7
Gambar Neural Network Multilayer



Sumber : Charu C Aggarwal (2015:4)

Jaringan yang ditampilkan pada Gambar 2.7 sepenuhnya terhubung, karena semua *node* pada satu lapisan terhubung ke semua *node* pada lapisan berikutnya. Selain itu, setiap koneksi *node* jaringan memiliki bobot yang terkait (tidak ditampilkan dalam diagram). Perhatikan bahwa *node* dalam lapisan arsitektur jaringan yang sama tidak terhubung satu sama lain.

Neural Network beroperasi dalam dua fase. Fase pertama disebut fase belajar. Selama pembelajaran jaringan, nilai *input* yang terkait dengan setiap *instance* masuk ke jaringan pada lapisan *input*. Satu *node input-layer* ada untuk setiap atribut *input* yang terkandung dalam data. Nilai *output* aktual untuk setiap *instance* dihitung dan dibandingkan dengan

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.





output jaringan yang diinginkan. Setiap kesalahan antara *output* yang diinginkan dan yang dikomputasi disebarakan kembali melalui jaringan dengan mengubah nilai-nilai koneksi-berat. Pelatihan berakhir setelah sejumlah iterasi atau ketika jaringan menyatu ke tingkat kesalahan minimum yang ditentukan sebelumnya. Selama fase operasi kedua, bobot jaringan ditetapkan, dan jaringan digunakan untuk mengklasifikasikan *instance* baru.

Menurut Stephan Kudeyba (2014:87) *Neural Network* awalnya dikembangkan oleh para peneliti yang mencoba untuk meniru neurofisiologi otak manusia. Dengan menggabungkan banyak elemen komputasi sederhana (*neuron* atau unit) ke dalam sistem yang sangat saling berhubungan, para peneliti ini berharap dapat menghasilkan fenomena kompleks seperti kecerdasan. Meskipun ada kontroversi tentang apakah jaringan saraf benar-benar cerdas, tidak bisa dipungkiri bahwa algoritma ini sangat kuat dalam mendeteksi hubungan nonlinier kompleks dalam data berdimensi tinggi.

Keuntungan utama dari jaringan saraf adalah mereka sangat kuat dalam memodelkan tren nonlinier. Mereka juga berguna ketika hubungan antara variabel input (termasuk interaksi) dipahami secara samar-samar. Anda juga dapat menggunakan jaringan syaraf tiruan untuk memenuhi kuadrat terkecil dan model regresi logistik dalam bentuk yang paling sederhana.

Jaringan syaraf juga sering disebut sebagai kotak hitam; meskipun mereka sangat pandai memprediksi, mereka sering sulit dipahami. Seperti analisis regresi, jaringan syaraf mengharuskan Anda untuk mengganti (menafsirkan) nilai yang hilang terlebih dahulu sebelum Anda melatih jaringan. Algoritma pembelajaran mesin paling populer saat ini adalah jaringan syaraf pembelajaran dalam. Jaringan neural generasi baru ini memiliki arsitektur yang sangat dalam dan kompleks yang menangani masalah minimum lokal dengan menambahkan lapisan tanpa pengawasan tanpa pengawasan dari arsitektur yang dalam.



M. Generalized Linear Model

Menurut Alan Agresti (2017:2) Model regresi linier biasa menggunakan linearitas untuk menggambarkan hubungan antara mean dari variabel respon dan satu set variabel penjelas, dengan kesimpulan mengasumsikan bahwa distribusi respons adalah normal.

Sedangkan pada *Generalized linear models (GLMs)* memperluas model regresi linier standar untuk mencakup variabel non-normal, distribusi respons dan kemungkinan fungsi nonlinear dari mean. Mereka memiliki tiga komponen.

1. Komponen acak: Ini menentukan variabel respon y dan distribusi probabilitasnya. Observasi $y = (y_1, \dots, y_n)^T$ pada distribusi tersebut diperlakukan sebagai independen.
2. Prediktor linier: Untuk vektor parameter $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$, dan matriks model X yang berisi nilai p variabel penjelas untuk observasi, prediktor linier adalah $X\beta$.
3. Fungsi tautan: Ini adalah fungsi yang diterapkan ke setiap komponen $E(y)$ yang menghubungkannya dengan prediksi linear, $g[E(y)] = X\beta$.

Menurut Menurut Jiawei Han, Micheline Kamber, dan Jian Pei (2012:599) *Generalized Linear Model* memungkinkan variabel respon kategori (nominal) (atau beberapa transformasi) untuk dikaitkan dengan satu set variabel prediktor dengan cara yang mirip dengan pemodelan variabel respon numerik menggunakan regresi linier. Model linear umum termasuk regresi logistik dan regresi Poisson.



N. Effect Size

Menurut *Wikipedia* dalam statistik, ukuran efek (*effect size*) adalah ukuran kuantitatif dari besarnya suatu fenomena. Contoh ukuran efek adalah korelasi antara dua variabel, koefisien regresi dalam regresi, perbedaan rata-rata, atau bahkan risiko di mana sesuatu terjadi seperti berapa banyak orang yang bertahan hidup setelah serangan jantung untuk setiap orang yang tidak bertahan hidup.

Menurut Mildred L. Patten dan Michelle Newhart (2018:265) ukuran efek (*effect size*) besarnya perbedaan (seperti ukuran) yang dinyatakan pada skala standar.

O. Tinjauan Studi

Dalam penelitian ini, penulis telah mengambil beberapa contoh referensi dari penelitian terdahulu untuk membantu menyelesaikan penelitian ini:

1. Penelitian “Prediksi Indeks Pembangunan Manusia (IPM) Berdasarkan Jumlah Pengguna Internet, Produk Domestik Regional Bruto (PDRB), Jumlah Pekerja, Jumlah Penduduk Miskin dan Populasi Penduduk Pada Tahun 2010” (Adit Dwi Fajaryanto dan Andry Alamsyah: 2016) yang dimuat pada Jurnal Telkom University (link http://repository.telkomuniversity.ac.id/pustaka/files/121508/jurnal_eproc/clusterin-g-data-indeks-pembangunan-manusia-dan-produk-domestik-bruto-untuk-indentifikasi-pemerataan-pembangunan-di-indonesia.pdf). Penelitian ini dilakukan untuk memprediksi nilai IPM berdasarkan jumlah pengguna internet, Produk Domestik Regional Bruto (PDRB), jumlah pekerja, jumlah penduduk miskin, dan populasi penduduk pada tahun 2010 dengan menggunakan algoritma *Decision Tree* dan *Artificial Neural Network*.

Hak cipta dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



- a. Penelitian ini menerapkan algoritma prediksi *Decision Tree* dan *Artificial Neural Network* untuk memprediksi nilai IPM berdasarkan 5 variabel prediktor yaitu jumlah pengguna internet, Produk Domestik Regional Bruto (PDRB), jumlah pekerja, jumlah penduduk miskin, dan populasi penduduk. Dari penelitian ini penulis mengambil beberapa referensi tentang cara mendapatkan sumber data sekunder yang akurat dari bps.go.id dan mengetahui cara kerja algoritma *Neural Network* untuk melakukan prediksi nilai IPM dengan tingkat akurasi yang terpercaya.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

2. Penelitian “*Data Mining* untuk Indeks Pembangunan Manusia” (Nabilah Izatani: 2010) yang dimuat dalam jurnal Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung (STEI ITB) (link jurnal: Institut Teknologi Bandung). Penelitian ini dilakukan untuk membahas tentang pemanfaatan ilmu statistik dalam metode penghitungan IPM lewat *Data Mining*.
 - a. Penelitian ini membahas bagaimana menerapkan ilmu statistik dalam *data mining* menggunakan sampel data IPM. Penulis mengambil referensi penggunaan rumus penghitungan IPM yang digunakan untuk menghitung IPM serta indikator-indikator penting dalam menentukan variabel yang terbaik untuk menghitung nilai IPM.

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.