



# KAJIAN KOMPARASI PENGUKURAN AKURASI ALGORITMA NEURAL NETWORK DAN GENERALIZED LINEAR MODEL UNTUK PREDIKSI DATA INDEKS PEMBANGUNAN MANUSIA DI INDONESIA

Ryan Nugraha <sup>1)</sup> dan Joko Susilo, S.Kom., M.M., M.Kom. <sup>2)</sup>

1) Alumni Program Studi Sistem Informatika  
2) Staff Pengajar Program Studi Sistem Informatika  
Institut Bisnis dan Informatika Kwik Kian Gie  
Jl. Yos Sudarso Kav.87 Sunter Jakarta Utara 14350  
<http://www.kwikkiangie.ac.id>  
[joko.susilo@kwikkiangie.ac.id](mailto:joko.susilo@kwikkiangie.ac.id)

## ABSTRAK

As time goes on, it cannot be denied that the role of information technology is inseparable in human life, especially in the digital era as it is now. The types of human needs that are diverse and continue to grow also contribute to the reasons why digital computing is becoming a growing thing at all times. Data processing technology is also demanded to be able to always adapt to offset human needs every day. The availability of a good data processing system will certainly be able to solve various problems and process information quickly and efficiently. One of the benefits that we can get by implementing information technology and a capable system is the ability to analyze and predict large amounts of data.

**Kata Kunci :** Prediction, Farmers Exchange Rate, Food Crop Farmers Exchange Rate, Data Mining, Linear Regression, Support Vector Machine, Rapidminer.

## 1. PENDAHULUAN

Seiring bertambahnya waktu, tidak dapat dipungkiri bahwa peran teknologi informasi menjadi hal yang tidak dapat dipisahkan dalam kehidupan manusia terutama di era digital seperti sekarang. Jenis kebutuhan manusia yang beragam dan terus bertambah juga turut menjadi alasan mengapa komputasi digital menjadi hal yang terus berkembang setiap saat. Teknologi pengolahan data juga turut dituntut agar dapat selalu beradaptasi dalam mengimbangi kebutuhan manusia setiap harinya. Hal ini bertujuan agar bisa menghasilkan sebuah sistem yang dapat memproses data dalam jumlah yang banyak dalam waktu singkat. Dalam kehidupan manusia setiap hal atau sebuah kejadian yang terjadi tiap waktu mampu diterjemahkan sebagai sebuah fakta. Kemudian fakta tersebut dapat diolah kembali menjadi sebuah data yang selanjutnya dapat diproses menjadi sebuah informasi yang memiliki pola atau bentuk. Pola dan bentuk yang terdapat dalam

informasi tersebut kemudian dapat diolah dalam sebuah sistem. Pengolahan informasi dalam jumlah yang besar tentunya membutuhkan sebuah sistem yang mampu melakukan proses komputasi dan pengolahan data yang akurat, cepat, serta efisien.

Berbagai macam jenis sistem yang telah berkembang dan hadir di kehidupan manusia sudah pasti bertujuan untuk mempercepat kinerja pengolahan informasi agar dapat menyelesaikan bermacam jenis masalah dalam kehidupan sehari-hari. Ketersediaan sistem pengolahan data yang baik tentunya akan mampu menyelesaikan berbagai masalah dan mengolah informasi dengan cepat serta efisien. Salah satu manfaat yang bisa kita peroleh dengan melakukan penerapan teknologi informasi dan sistem yang mumpuni adalah kemampuan untuk melakukan analisis dan prediksi data dalam jumlah yang sangat besar. Sejumlah data dengan volume yang besar dapat di analisis dan diterjemahkan menjadi sebuah pola agar proses pengukuran dan prediksi nilai sesuai dengan



kebutuhan di masa depan. Berdasarkan alasan tersebut, penulis kemudian memutuskan untuk melakukan penelitian untuk mengukur tingkat akurasi algoritma prediksi dengan sampel data Indeks Pembangunan Manusia (IPM) per provinsi di Indonesia.

Indeks Pembangunan Manusia (IPM) atau Human Development Index (HDI) adalah sebuah pengukuran dan perbandingan dari angka harapan hidup, angka melek huruf, tingkat pendidikan, dan standar kehidupan yang berlaku untuk semua negara di seluruh dunia. Nilai dari IPM ini kemudian digunakan untuk menentukan klasifikasi apakah sebuah negara bisa dikategorikan sebagai negara maju, negara berkembang, atau negara terbelakang. IPM juga bisa digunakan untuk mengukur pengaruh dari kebijakan ekonomi terhadap kualitas hidup manusia di sebuah negara. Menurut *Wikipedia*, penghitungan indeks ini dikembangkan oleh pemenang nobel India Amartya Sen dan seorang ekonom Pakistan Mahbub ul Haq, serta dibantu oleh Gustav Ranis dari Yale University dan Lord Meghnad Desai dari London School of Economics. Sejak saat itu indeks ini kemudian digunakan oleh PBB dalam program pembangunan mereka setiap tahun.

Seperti digambarkan oleh Amartya Sen. Indeks ini bisa disebut sebagai pengukuran vulgar karena berbagai batasan yang muncul. IPM lebih berfokus pada hal-hal yang lebih sensitif, dan berguna dibandingkan hanya sekadar mengukur pendapat perkapita yang selama ini sering digunakan. IPM juga bisa menjadi sebuah jembatan untuk para peneliti yang lebih serius dalam meneliti hal-hal yang lebih terperinci dalam pembuatan laporan pembangunan manusia. Dalam IPM, pencapaian rata-rata sebuah negara bisa diukur dalam 3 dimensi dasar yaitu tingkat kehidupan yang sehat dan panjang umur yang diukur lewat angka harapan hidup saat lahir, tingkat pengetahuan yang diukur lewat tingkat baca tulis orang dewasa, dan standar kelayakan hidup yang diukur lewat logaritma natural dari *Produk Domestik Bruto (PDB)* per kapita dalam partisipasi daya beli masyarakat.

Indonesia sebagai negara yang memiliki jumlah penduduk yang besar di dunia

tentunya juga tidak terlepas dari pengukuran IPM dari PBB. Data IPM di Indonesia dikumpulkan dan dihitung melalui Badan Pusat Statistik (BPS) yaitu sebuah lembaga pemerintah non kementerian yang bertanggung jawab langsung kepada presiden untuk melakukan pengumpulan data tentang berbagai jenis bidang di Indonesia untuk kebutuhan pemerintah dan publik. Data mengenai IPM yang dikumpulkan tersebut meliputi data angka harapa hidup, angka melek huruf, lama sekolah serta pengeluaran perkapita. Data-data tersebut dikumpulkan berdasarkan setiap provinsi di Indonesia setiap tahun yang kemudian digunakan untuk menghitung IPM.

Berbagai data yang kemudian telah dikumpulkan oleh BPS untuk IPM dapat diakses melalui situs resmi BPS yaitu <http://bps.go.id> atau melalui portal data publik di Indonesia yaitu <http://data.go.id>. Data indeks IPM yang dikumpulkan oleh BPS memiliki tingkat akurasi yang tinggi karena dikumpulkan berdasarkan data tiap provinsi yang dilakukan dengan melakukan observasi langsung ke lapangan lewat tenaga ahli statistik dalam proses sehingga data tersebut sangat layak untuk digunakan dalam objek penelitian.

Lewat metode data mining, data IPM di Indonesia dapat diprediksi dengan menerapkan algoritma prediksi *Neural Network* dan *Generalized Linear Model*. Dengan menerapkan algoritma ini, kita dapat melakukan prediksi nilai IPM di masa depan secara sistematis dan cepat tanpa harus mengumpulkan variabel terkait seperti angka harapa hidup, angka melek huruf, lama sekolah atau pengeluaran perkapita yang membutuhkan waktu yang cukup lama. Dalam penelitian ini penulis juga akan melakukan pengukuran kelayakan kedua algoritma tersebut agar mampu memilih algoritma prediksi yang paling tepat untuk mendapatkan data IPM di masa depan berdasarkan data pembanding dan sampel data yang telah disiapkan dari 2009 sampai dengan 2012.

Proses pengolahan data IPM secara manual tentu saja tidak menutup adanya berbagai masalah yang timbul. Penggunaan metode konvensional untuk pengumpulan data IPM secara manual yang memakan waktu sehingga data prediksi IPM tidak

© Hak Cipta Dilindungi Undang-Undang  
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.



dapat diperoleh dengan cepat karena harus menunggu proses survey dan pendataan secara berkala yang kurang efektif. Pengumpulan data yang memakan waktu lama ini juga menyebabkan belum ditentukannya data acuan (*training*) yang tepat pada periode tertentu untuk pemodelan prediksi menggunakan metode data mining. Tanpa data acuan (*training*) tentunya akurasi algoritma *prediksi Neural Network* dan *Generalized Linear Model* yang belum terukur masih belum diketahui kelayakannya untuk memprediksi data IPM. Selain itu, pengujian kelayakan algoritma tersebut tentu membutuhkan pembuktian jumlah variabel yang paling tepat untuk proses pengetesan. Karena perbandingan jumlah variabel untuk prediksi yang belum dilakukan sehingga tidak bisa mengetahui jumlah variabel yang paling tepat untuk menentukan akurasi algoritma *prediksi Neural Network* dan *Generalized Linear Model*. Hal ini juga berpengaruh karena laporan penghitungan akurasi algoritma prediksi berdasarkan variabel dan periode yang belum tercatat sehingga dokumentasi pengujian belum bisa dilakukan.

Berdasarkan hal-hal diatas, penulis kemudian memutuskan untuk melakukan penelitian ini yang bertujuan untuk melakukan pengujian dan pengukuran akurasi algoritma *Neural Network* dan *Generalized Linear Model* dengan variabel dan data periode masa lalu agar dapat melakukan prediksi nilai IPM secara akurat, efisien, dan mendekati data asli di masa depan.

## 2. LANDASAN TEORI

### 2.1 Sistem Informasi

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 [8]. Pendapat lain mengatakan bahwa sistem informasi adalah seperangkat komponen yang saling terkait yang mengumpulkan, memproses, menyimpan, dan menyebarluaskan data dan informasi [9].

### 2.2 Data

Data adalah sebuah fakta mentah [9] atau aliran fakta mentah yang mewakili peristiwa yang terjadi dalam organisasi atau lingkungan fisik sebelum diatur dan disusun menjadi bentuk yang dapat dipahami dan digunakan [8].

### 2.3 Informasi

Menurut [9] mengartikan bahwa Informasi adalah kumpulan data yang diatur dan diproses sehingga memiliki nilai tambahan di luar nilai fakta individu.

### 2.4 Database

*Database* merupakan kumpulan terorganisir dari fakta-fakta dan informasi, biasanya terdiri dari dua atau lebih terkait file data [9].

### 2.4 Database Management Systems

*Database Management Systems* adalah aplikasi perangkat lunak yang digunakan untuk membuat, menyimpan, mengatur, dan mengambil data dari database tunggal atau beberapa basis data [12].

### 2.5 Data Warehouse

Menurut [13] sebagai pusat penyimpanan data yang berisi informasi yang diambil dari berbagai sumber yang dapat digunakan untuk analisis, pengumpulan intelijen, dan perencanaan strategis.

### 2.6 Data Mining

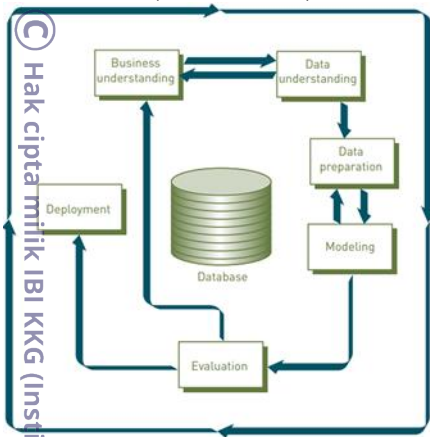
Menurut [4] *data mining* adalah proses menemukan pola yang menarik dan pengetahuan dari sejumlah besar data atau data mining sebagai proses penemuan struktur yang menarik dalam data [11]

### 2.7 Tahap-tahap Data Mining

Pendekatan terstruktur untuk perencanaan dan pelaksanaan proyek penambangan data yang disebut dengan *The Cross-Industry Process for Data Mining (CRISP-DM)* [9]. Ada 6 tahapan antara lain:



**Gambar 1. Tahapan data mining (CRISP-DM)**

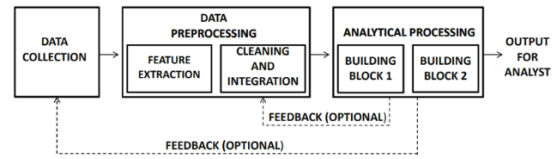


Hak Cipta Dilindungi Undang-Undang

1. **Business Understanding/Organizational Understanding** (Pemahaman bisnis/organisasi) : 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 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) : Terapkan model ke dalam proses pengambilan keputusan.

Institut Bisnis dan Informatika Kwik Kian Gie

Menurut [2] dibagi dalam dalam 3 proses besar yaitu:



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.
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.
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.

### 2.8 Metode Dalam Data Mining

Menurut [4]. ada beberapa jenis fungsionalitas database yang dapat yang dapat dilakukan dalam *data mining* antara lain adalah :

1. **Class/Concept Characterization and Discrimination** (Deskripsi kelas/konsep : karakterisasi dan diskriminasi) : Entri data yang dapat dikaitkan dengan kelas atau konsep.
2. **Mining Frequent Patterns Associations, and Correlations** (Penggalian pola yang sering muncul : asosiasi dan korelasi) : Meneliti pola yang sering terjadi di data

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.



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.

**2.9 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.

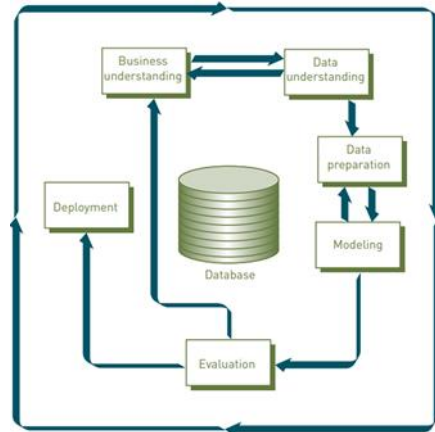
Menurut Badan Pusat Statistik (BPS) IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya.

**3. METODOLOGI PENELITIAN**

**3.1 Teknik Analisa Data**

Penulis menggunakan teknik analisa data berdasarkan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yaitu sebuah metode data mining 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 3. Data mining cycle (CRISP-DM)**



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.

Hak Cipta Dilindungi Undang-Undang

Hak Cipta milik IBI KKG (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.



6. Deployment (penyebaran):  
erapkan model ke dalam proses pengambilan keputusan.

### 3.2 Teknik Pengukuran

#### Pengukuran Akurasi Metode Prediksi :

Pengukuran kelayakan metode prediksi penulis menggunakan rumus *Root Mean Squared Error* untuk menentukan tingkat kesalahan prediksi data berdasarkan pengaruh dari tiap variabel :

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y)^2}{n}}$$

$y_t$  = Data Prediksi  
 $y$  = Data Aktual  
 $n$  = Jumlah Data

#### Pengukuran Penyimpangan Prediksi :

Untuk mengukur penyimpangan data prediksi penulis melakukan pencarian rata-rata kesalahan (*Mean Error*), rata-rata kesalahan absolut (*Mean Absolute Error*), rata-rata persentase kesalahan (*Mean Percentage Error*), dan rata-rata persentase kesalahan absolut (*Mean Absolute Percentage Error*).

*Mean Error* (ME) :

$$ME = \frac{\sum E}{n}$$

$E$  = Data Aktual – Data Prediksi  
 $n$  = Jumlah Data

*Mean Absolute Error* (MAE) :

$$MAE = \left| \frac{\sum E}{n} \right|$$

$E$  = Data Aktual – Data Prediksi  
 $n$  = Jumlah Data

*Mean Percentage Error* (MPE) :

$$MPE = \frac{\sum \left( \frac{E}{A} \right) \times 100\%}{n}$$

$E$  = Data Aktual – Data Prediksi  
 $n$  = Jumlah Data  
 $A$  = Data Aktual

*Mean Absolute Percentage Error* (MAPE) :

$$MAPE = \left| \frac{\sum \left( \frac{E}{A} \right) \times 100\%}{n} \right|$$

$E$  = Data Aktual – Data Prediksi  
 $n$  = Jumlah Data  
 $A$  = Data Aktual

### 3.3 Teknik Perancangan GUI

Dalam penelitian ini, laporan penelitian akan didokumentasikan dan disajikan dalam bentuk *Graphical User Interface (GUI)*. Perancangan GUI ini akan dilakukan pada aplikasi Microsoft Access. Sebelum proses perancangan dilakukan, hasil prediksi dan pengukuran akurasi yang didapat dari algoritma *Neural Network* dan *Generalized Linear Model* akan di export ke dalam *Microsoft Access 2013* agar bisa disajikan dalam bentuk tabel yang lebih mudah dipahami dan bisa melakukan visualisasi data.

Hasil pengujian data IPM akan dilakukan dari tahun 2009-2012 yang kemudian diambil sebagai data training. Data training ini kemudian akan dijadikan model untuk penerapan algoritma *Neural Network* dan *Generalized Linear Model* yang kemudian akurasi akan diukur berdasarkan angka.

Dari hasil pengujian tersebut pembagian data akan dilakukan yang terdiri dari data tahun 2009-2012 sebagai data training untuk dijadikan model algoritma yang ditentukan, yang kemudian akan diuji kelayakannya

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.

Kwik Kian Gie Institut Bisnis dan Informatika



berdasarkan angka *Root Mean Squared Error (RMSE)* untuk mendapatkan angka tingkat kelayakan masing-masing algoritma yang digunakan. Kemudian, untuk hasil data prediksi dari kedua algoritma tersebut akan diukur dengan metode pengukuran penyimpangan prediksi untuk menentukan hasil pengukuran algoritma yang paling akurat.

Penulis kemudian akan melakukan import database dan mengimplementasikan hasil prediksi tersebut kedalam Microsoft Access 2013 untuk kemudian melakukan perancangan *GUI* sederhana yang dilakukan untuk melakukan proses penghitungan penyimpangan keakuratan prediksi dan menghasilkan laporan komparasi data prediksi kepada para pembaca.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Rancangan Sistem

Perancangan sistem yang dilakukan oleh penulis pada penelitian kali ini dilakukan dengan membuat sebuah *Graphical User Interface (GUI)* yang diperuntukkan untuk para pembaca dan pengguna aplikasi. Dengan adanya *GUI* yang dibuat melalui *Microsoft Access 2013*, dapat menyediakan laporan penelitian yang tepat dalam pengujian algoritma *Neural Network* dan *Generalized Linear Model* dalam mengukur akurasi prediksi nilai Indeks Pembangunan Manusia (IPM) di Indonesia. Dalam rancangan *GUI* ini juga disertakan hasil error margin dan grafik tingkat akurasi masing-masing algoritma dalam memprediksi nilai IPM.

### 4.2 Rancangan Basis Data

Pada Penelitian kali ini, data Indeks Pembangunan Manusia (IPM) yang diambil dan diunduh dari <http://data.go.id/> akan diolah menjadi bentuk yang bisa diolah dalam aplikasi Rapidminer. Proses pengolahan data ini meliputi konversi data training dari format .csv menjadi .xlsx, kemudian akan diambil variabel-variabel penting yang digunakan untuk proses pengetesan. Variabel data yang diambil antara lain adalah angka harapan hidup, angka melek huruf, lama sekolah, pengeluaran perkapita, dan Nilai IPM.

Setelah data training sudah siap, langkah selanjutnya adalah melakukan pengukuran kelayakan algoritma *Neural Network* dan *Generalized Linear Model* berdasarkan nilai standar deviasi (*Root Mean Squared Error*). Setelah proses pengukuran standar deviasi dilakukan, akan dilakukan proses pemodelan prediksi untuk mendapatkan nilai akurasi tiap-tiap algoritma. Data nilai akurasi tersebut kemudian akan dimasukkan ke dalam file terpisah melalui *Microsoft Excel* yang kemudian akan di export ke dalam aplikasi *Microsoft Access*. Kemudian perancangan *GUI* akan dilakukan untuk membentuk modul-modul laporan yang dibutuhkan untuk para pembaca dalam menilai hasil penelitian ini. Tahap rancangan basis data akan dilakukan melalui beberapa tahap yaitu::

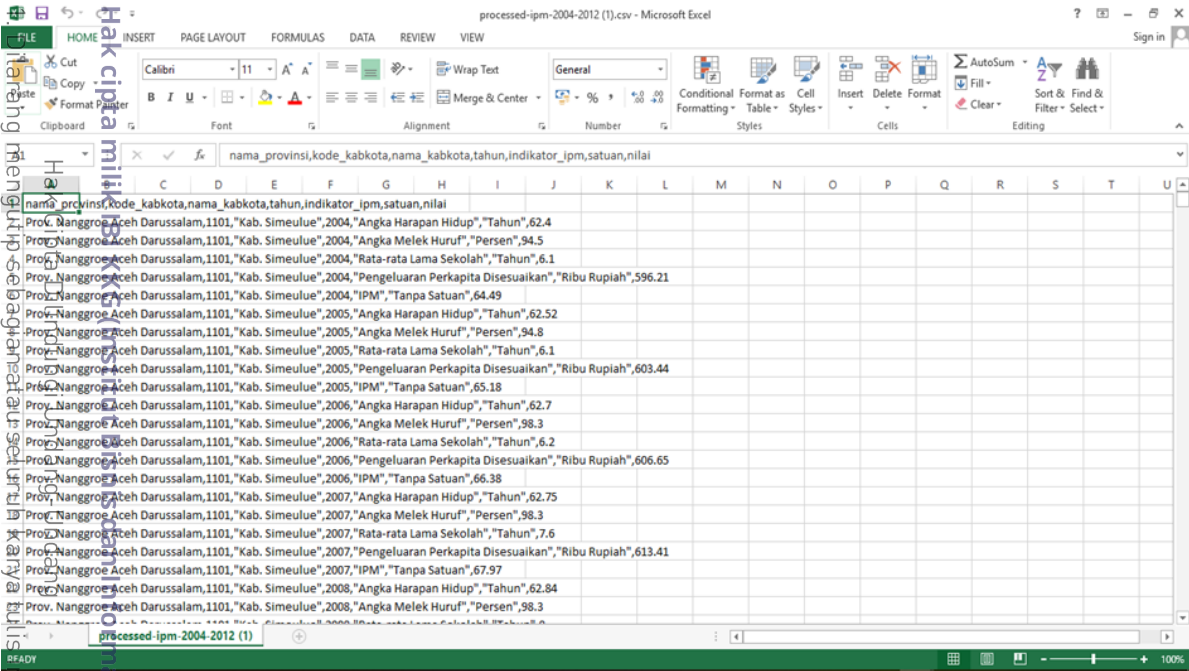
1. Persiapan Data Prediksi (*data Mining*).
2. Pengujian Kelayakan Algoritma Prediksi.
3. Pemodelan Algoritma Prediksi.

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.

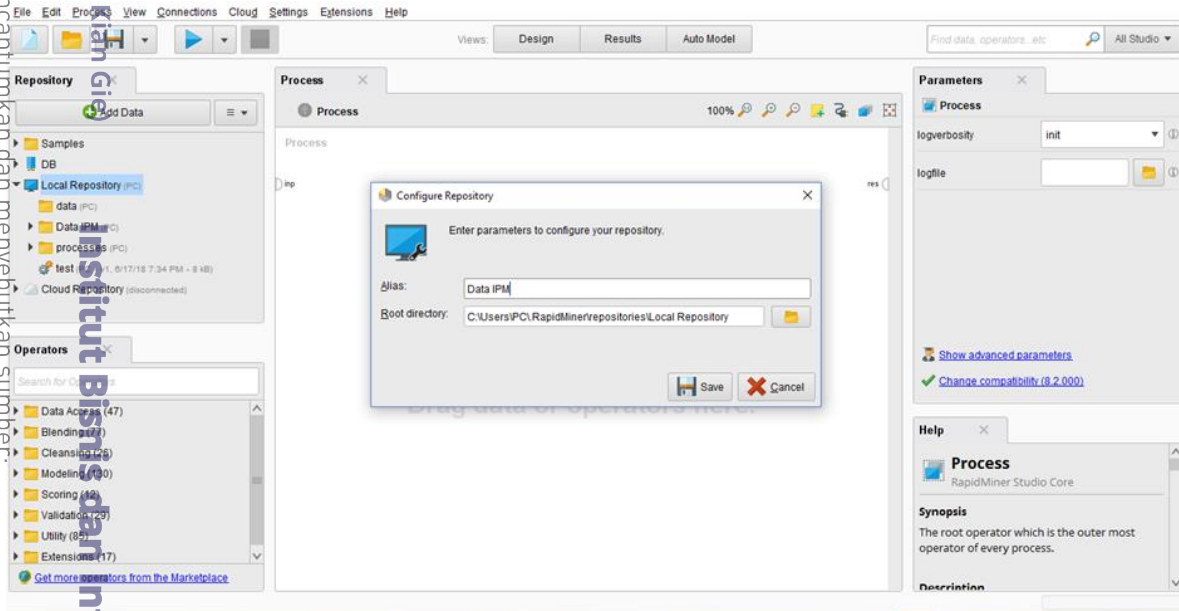
b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.  
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Gambar 4.1  
Persiapan Data



Gambar 4.2  
Repository Data IPM



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

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

b. Penugutan tidak merugikan kepentingan yang wajar IBIKKG.

Hak cipta milik Institut Bisnis dan Informatika Kwik Kian Gie

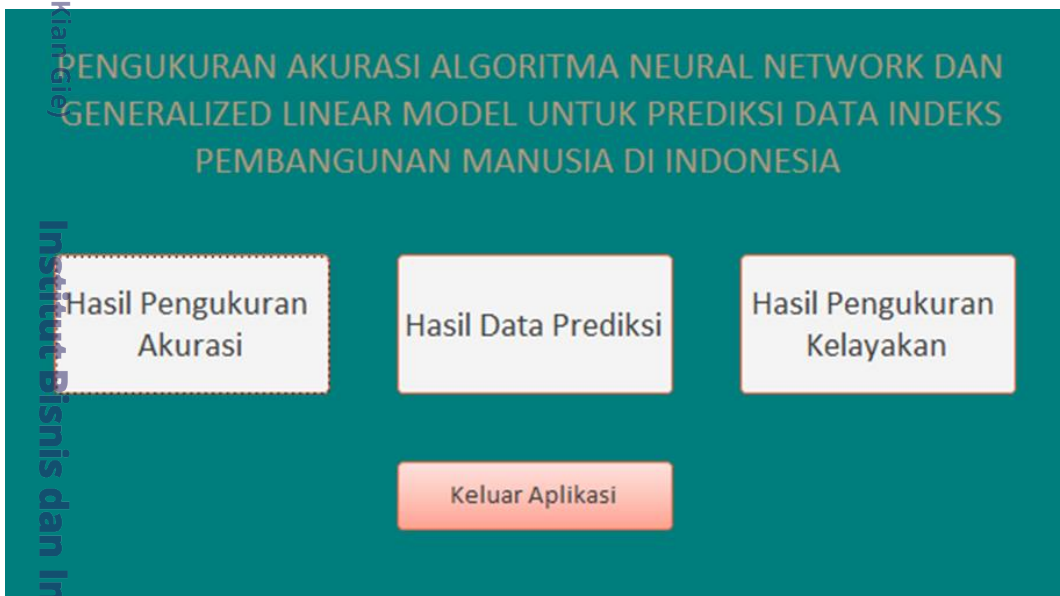




**Tabel 4.1**  
**Perbandingan *Error Margin 2* Algoritma Prediksi**

Error Margin Algoritma Neural Network dan GLM					
Tahun	Jumlah Variabel	Variabel	MAE	MAPE (dalam %)	Tingkat Akurasi (dalam %)
GLM 1 Tahun (2011)	4	Semua	0.013077	0.019792133	99.980207867
GLM 2 Tahun (2010-2011)	4	Semua	0.013343	0.020132933	99.979867067
GLM 3 tahun (2009-2011)	4	Semua	0.013767	0.020723355	99.979276645
NN 2 Tahun (2010-2011)	4	Semua	0.027519	0.040175142	99.959824858
NN 1 Tahun (2011)	4	Semua	0.029545	0.042886348	99.957113652
NN 3 tahun (2009-2011)	4	Semua	0.030411	0.043790782	99.956209218

**Gambar 4.3**  
**Tampilan Menu Utama**

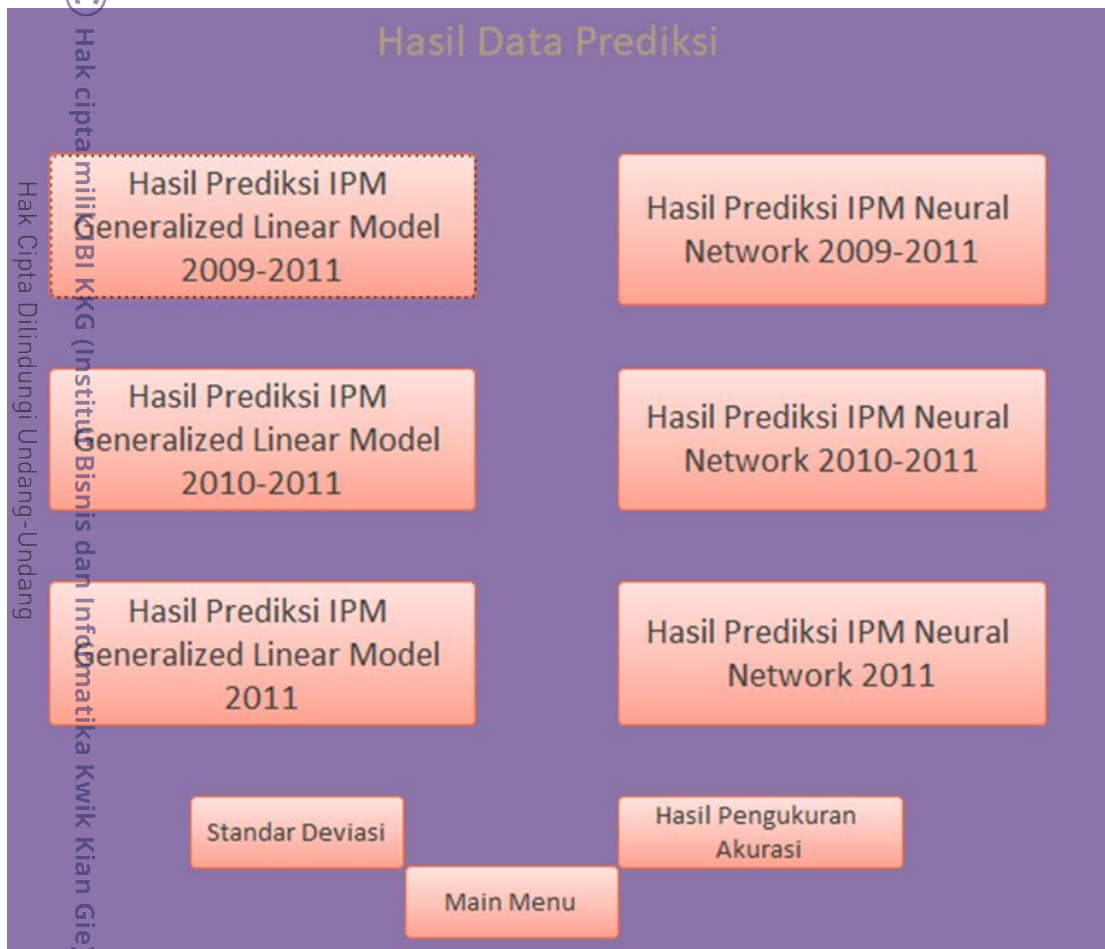


© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

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.

**Gambar 4.4**  
**Tampilan Form Hasil Data Prediksi**



**Gambar 4.5**  
**Tampilan Report Hasil Data Prediksi**

Hasil Prediksi IPM Tahun 2012 (Generalized Linear Model 2009-2011)

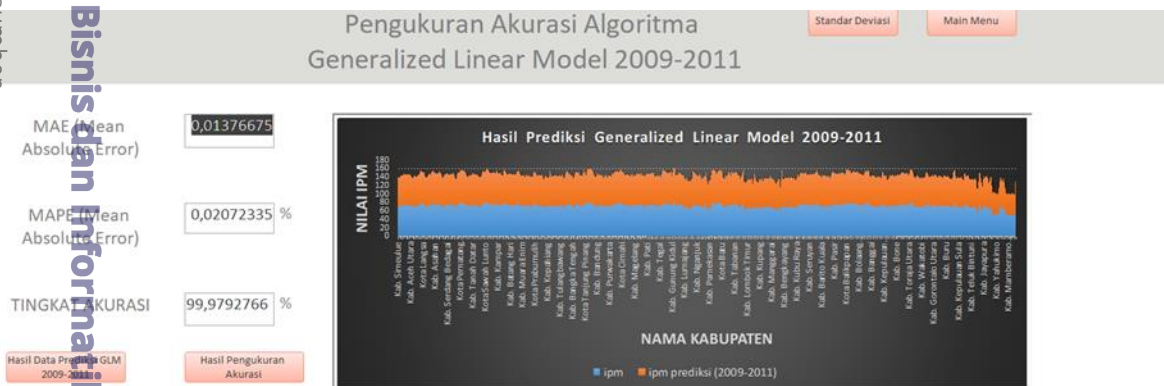
kode_kabkota	nama_kabkota	ipm	ipm prediksi (2009-2011)	Deviasi Absolut	Nilai Bagi Absolut	Hasil Pengukuran Akurasi GUM 2009-2011	Hasil Data Prediksi
1101	Kab. Simeulue	70,0928	70,1064	0,0136368	0,0001946		
1102	Kab. Aceh Singkil	69,3657	69,3716	0,0059065	0,0000852		
1103	Kab. Aceh Selatan	70,7067	70,7115	0,0047445	0,0000671		
1104	Kab. Aceh Tenggara	72,2477	72,2516	0,0039034	0,0000540		
1105	Kab. Aceh Timur	71,1716	71,1687	0,0029483	0,0000414		
1106	Kab. Aceh Tengah	74,4246	74,4232	0,0013946	0,0000187		
1107	Kab. Aceh Barat	71,7286	71,7295	0,0009684	0,0000135		
1108	Kab. Aceh Besar	74,1287	74,1294	0,0006692	0,0000090		
1109	Kab. Pidie	72,8127	72,8085	0,0042292	0,0000581		
1110	Kab. Bireuen	73,7042	73,6957	0,0084903	0,0001152		
1111	Kab. Aceh Utara	73,0689	73,0672	0,0016992	0,0000233		
1112	Kab. Aceh Barat Daya	71,5258	71,5252	0,0005938	0,0000083		
1113	Kab. Gayo Lues	68,5405	68,5649	0,0243425	0,0003552		
1114	Kab. Aceh Tamiang	71,6539	71,6553	0,0013717	0,0000191		
1115	Kab. Nagan Raya	70,6362	70,6376	0,0013624	0,0000193		
1116	Kab. Aceh Jaya	70,3476	70,3567	0,0090846	0,0001291		
1117	Kab. Bener Meriah	71,8640	71,8654	0,0013955	0,0000194		
1118	Kab. Pidie Jaya	73,1295	73,1255	0,0039896	0,0000546		
1171	Kota Banda Aceh	78,4958	78,5045	0,0086841	0,0001106		
1172	Kota Sabang	76,8829	76,8788	0,0040089	0,0000521		
1173	Kota Langsa	74,7455	74,7496	0,0041443	0,0000554		
1174	Kota Lhokseumawe	77,2307	77,2228	0,0079569	0,0001030		

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 4.82  
Tampilan Form Hasil Pengukuran Akurasi



Gambar 4.83  
Contoh Tampilan Report Pengukuran Akurasi



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 menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



**Gambar 4.84**  
**Tampilan Report Nilai RMSE**

Tahun	Jumlah Variabel	Variabel	RMSE (Mikro)
GLM 3 tahun (2009-2011)	4	Semua	0,0200
GLM 2 Tahun (2010-2011)	4	Semua	0,0200
GLM 1 Tahun (2011)	4	Semua	0,0200
NN 3 tahun (2009-2011)	4	Semua	0,0480
NN 2 Tahun (2010-2011)	4	Semua	0,0490
NN 1 Tahun (2011)	4	Semua	0,0560
ipm 3 tahun (2009-2011)	3	ipm (AHH/AMF/PP)	0,6290
ipm 2 tahun (2009-2011)	3	ipm (AHH/AMF/PP)	0,6300
ipm 1 tahun (2009-2011)	3	ipm (AHH/AMF/PP)	0,6590
GLM2 tahun (2009-2011)	3	ipm (AHH/AMF/PP)	0,7100
GLM3 tahun (2009-2011)	3	ipm (AHH/AMF/PP)	0,7110
GLM 1 tahun (2009-2011)	3	ipm (AHH/AMF/PP)	0,7190
NN 2 tahun (2009-2011)	3	ipm (AHH/LS/PP)	1,0220
ipm 3 tahun (2009-2011)	3	ipm (AHH/LS/PP)	1,0340
NN 1 tahun (2009-2011)	3	ipm (AHH/LS/PP)	1,1390
GLM2 tahun (2009-2011)	3	ipm (AMF/LS/PP)	1,3350
GLM 1 tahun (2009-2011)	3	ipm (AMF/LS/PP)	1,3380
GLM 3 tahun (2009-2011)	3	ipm (AMF/LS/PP)	1,3390
GLM 3 tahun (2009-2011)	3	ipm (AHH/AMF/LS)	1,3570
GLM 1 tahun (2009-2011)	3	ipm (AHH/AMF/LS)	1,3640
GLM 2 tahun (2009-2011)	3	ipm (AHH/AMF/LS)	1,3650
NN 1 tahun (2009-2011)	3	ipm (AMF/LS/PP)	1,4460
ipm 3 tahun (2009-2011)	3	ipm (AMF/LS/PP)	1,4520
NN 2 tahun (2009-2011)	3	ipm (AHH/AMF/LS)	1,4960
NN 2 tahun (2009-2011)	3	ipm (AMF/LS/PP)	1,5800

## 5 SIMPULAN

Berdasarkan penelitian Pengukuran Akurasi Algoritma *Neural Network* dan *Generalized Linear Model* untuk Prediksi Data Indeks Pembangunan Manusia di Indonesia penulis dapat menyimpulkan hal-hal berikut :

1. Waktu pengumpulan data IPM yang dapat dipersingkat dengan menggunakan proses data mining tanpa menggunakan pendataan dan penghitungan konvensional yang memakan banyak waktu dan kurang efisien
2. Menentukan data acuan (*testing*) yang tepat untuk melakukan pemodelan prediksi yaitu data 1 tahun (2009-2011) untuk algoritma *Generalized Linear Model* dan data 2 tahun untuk algoritma *Neural Network*. Tercatat juga untuk mendapatkan hasil data IPM yang

paling akurat kita bisa menggunakan algoritma *Generalized Linear Model*.

3. Mendapatkan data perbandingan jumlah variabel untuk menentukan akurasi algoritma prediksi yaitu dengan menggunakan 4 variabel pada algoritma *Neural Network* dan *Generalized Linear Model*, dibandingkan dengan 3 variabel karena hasil akurasi yang semakin menurun. Dapat disimpulkan juga semakin banyak variabel pembanding, maka data IPM hasil pemodelan prediksi akan semakin akurat.
4. Terbuatnya laporan penghitungan akurasi algoritma prediksi *Neural Network* dan *Generalized Linear Model* lewat rancangan *Graphical User Interface (GUI)* pada aplikasi *Microsoft Access* untuk dokumentasi pengujian yang bisa

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.



kembali sebagai acuan atau referensi penelitian di masa depan.

## 6 REKOMENDASI

Dalam pembuatan penelitian karya akhir ini penulis sekiranya memberi saran lanjutan untuk para pembaca untuk dipertimbangkan antara lain :

1. Penggunaan aplikasi *Rapidminer* aplikasi pengolah data masih terbatas pada jumlah data yang tidak terlalu banyak (dibawah 10000 row), sehingga tidak mampu digunakan untuk data yang berjumlah lebih dari itu. Untuk versi aplikasi *Rapidminer* yang digunakan penulis bersifat gratis atau tidak berbayar, sehingga penulis berharap untuk para peneliti selanjutnya bisa menggunakan versi berbayar atau aplikasi sejenis lain yang mungkin menghasilkan data prediksi yang lebih akurat serta fitur analisa yang lebih baik.

2. Data IPM yang digunakan penulis sebagai data sampel prediksi yang dari website data.go.id masih tergolong sebagai data lama karena hanya sampai pada tahun 2012 dan masih menggunakan metode penghitungan IPM dengan rumus lama. Penulis berharap agar pada penelitian selanjutnya bisa digunakan data IPM terbaru yang disertai dengan model penghitungan baru agar hasil penelitian bisa teruji dengan lebih efektif.

3. Penggunaan algoritma *Neural Network* dan *Generalized Lunear Model* cukup sulit jika dilakukan secara manual karena penulis masih bergantung pada mesin pengolah data pada aplikasi *Rapidminer*,

Maka pada penelitian selanjutnya yang menggunakan algoritma yang sama disarankan untuk mencantumkan penghitungan secara manual sebagai referensi hasil prediksi yang lebih akurat.efektif.

4. Rancangan *GUI* yang dibuat oleh penulis sengaja dibuat secara simpel agar mudah dimengerti dan tidak memakan waktu karena implementasi algoritma dilakukan dengan segera, maka bila dirasa kurang memuaskan dapat dikembangkan atau ditambahkan sesuai keinginan pengembang selanjutnya agar tampilan *GUI* dapat lebih diperbaiki.
5. Sebagaimana sebuah penelitan, metode dan hasil yang diperoleh dari penelitian ini tidaklah menjadi sebuah hal yang "baku", karena itu hasil penelitan ini dapat saja diperbaharui di masa depan oleh peneliti selanjutnya agar dapat menemukan metode dan hasil yang lebih efektif.

## 7 DAFTAR REFERENSI

- [1] A.North, Agresti, Alan (2017), *Foundations of Linear and Generalized Linear Models*, New Jersey: John Wiley & Sons, Inc.
- [2] Anggrawal, Charu C (2015), *Data Mining The Textbook*, New York: Springer International Publishing.
- [3] Fajaryanto, Adit Dwi dan Andry Alamsyah (2016), *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*, Telkom University, Jawa Barat.



Hak Cipta Dilindungi Undang-Undang

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

**Institut Bisnis dan Informatika Kwik Kian Gie**

- [4] Han, Jiawei et al (2012), Data Mining Concepts and Techniques, Edisi ke-3, Waltham: Elsevier Inc.
- [5] Izatani, Nabilah (2010), Data Mining untuk Indeks Pembangunan Manusia, Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung, Bandung
- [6] Kroenke David M. (2014), Experienceing MIS, Edisi ke-5, New Jersey : Pearson Education, Inc.
- [7] Kudeyba, Stephan (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Boca Raton: Taylor and Francis Group. LLC.
- [8] Laudon, Kenneth C, dan Jane P. Laudon, (2017), Management Information Systems:Managing the Digital Firm, Edisi ke-15, New Jersey : Pearson Prentice Hall.
- [9] M.Stair, Ralph dan George W. Reynolds (2017), Principles of Information Systems, Edisi ke-13, Boston: Cengage Learning.
- [10] Patten, L. Mildred dan Michelle Newhart (2018), Understanding Research Methods An Overview of the Essentials, Edisi ke-10, New York: Routledge.
- [11] Roiger, J., Richard (2016) Data Mining A Tutorial-Based Primer, Edisi ke-2, Boca Raton: Taylor and Francis Group. LLC.
- [12] Valacich, Joseph dan Christoph Schneider (2017), Information Systems Today: Managing the Digital World, Edisi ke-8, New York: Pearson Prentice Hall.
- [13] Wallace, Patricia (2015), Introduction to Information System, Edisi ke-2, New Jersey: Perason Prentice Hall.

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 menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.