

**Pengenalan Emosi Menggunakan Pendekatan  
Pembelajaran Dalam Untuk Inteligensi Bisnis and  
Analitis**

**ALBERT EINSTEIN**



**SEKOLAH PASCASARJANA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2017**

**PERNYATAAN MENGENAI TESIS DAN  
SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA**

Dengan ini saya menyatakan bahwa tesis berjudul Pengenalan Emosi menggunakan Teknik Pembelajaran Dalam untuk Inteligensi Bisnis dan Analitis adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2017

Albert Einstein  
G651150211

## **RINGKASAN**

....

**SUMMARY**

....

**© Hak Cipta Milik IPB, Tahun 2017**

**Hak Cipta Dilindungi Undang-Undang**

*Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah; dan pengutipan tersebut tidak merugikan kepentingan IPB*

*Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB*

**Pengenalan Emosi Menggunakan Pendekatan  
Pembelajaran Dalam Untuk Inteligensi Bisnis dan  
Analitis**

**ALBERT EINSTEIN**

Tesis  
sebagai salah satu syarat untuk memperoleh gelar  
Magister Komputer  
pada  
Program Studi Ilmu Komputer

**SEKOLAH PASCASARJANA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2017**

Penguji Luar Komisi pada Ujian Tesis: ....

Judul: Pengenalan Emosi Menggunakan Pendekatan Pembelajaran Dalam untuk Intelligensi Bisnis dan Analitis

Nama: Albert Einstein

NIM: G651150211

Disetujui oleh  
Komisi Pembimbing

Dr. Eng. Taufik Djatna, S.TP, M.Si  
Ketua

Dr. Ir. Agus Buono, M.Si, M.Kom  
Anggota

Diketahui oleh

Ketua Program Studi  
Ilmu Komputer

Dekan Sekolah Pascasarjana

Dr. Ir. Agus Buono, M.Si, M.Kom

Tanggal Ujian: ....

Tanggal Lulus: ....



## **PRAKATA**

....

Bogor, Juli 2017

Albert Einstein

# DAFTAR ISI

|   |           |
|---|-----------|
| <b>PENDAHULUAN</b>                                  | <b>1</b>  |
| Latar Belakang . . . . .                            | 1         |
| Perumusan Masalah . . . . .                         | 2         |
| Tujuan Penelitian . . . . .                         | 2         |
| Manfaat Penelitian . . . . .                        | 2         |
| Ruang Lingkup Penelitian . . . . .                  | 3         |
| <b>TINJAUAN PUSTAKA</b>                             | <b>4</b>  |
| Pengenalan Emosi . . . . .                          | 4         |
| Representasi . . . . .                              | 6         |
| <i>Artificial Neural Networks</i> . . . . .         | 6         |
| <i>Convolutional Neural Networks</i> . . . . .      | 7         |
| <i>Stacked Convolutional AutoEncoder</i> . . . . .  | 9         |
| <i>Convolutional Deep Belief Networks</i> . . . . . | 10        |
| <b>METODE</b>                                       | <b>13</b> |
| Tahap Penelitian . . . . .                          | 13        |
| Alat dan Data . . . . .                             | 14        |
| Persiapan . . . . .                                 | 14        |
| Praproses . . . . .                                 | 15        |
| Pembelajaran . . . . .                              | 15        |
| Pengujian . . . . .                                 | 17        |
| <b>HASIL DAN PEMBAHASAN</b>                         | <b>19</b> |
| MNIST . . . . .                                     | 19        |
| Pembahasan . . . . .                                | 19        |
| <b>RENCANA DAN PERKEMBANGAN</b>                     | <b>20</b> |
| <b>Bibliografi</b>                                  | <b>21</b> |
| <b>A Log Keluaran</b>                               | <b>26</b> |
| MNIST . . . . .                                     | 26        |

## DAFTAR TABEL

|    |  |    |
|----|--|----|
| .1 | Arsitektur CNN beserta tipe setiap <i>layer</i> -nya (cont. MNIST, 28x28 piksel) | 16 |
| .2 | .....  | 17 |
| .1 | Tabel akurasi pengklasifikasi terhadap data set . . . . .                        | 19 |
| .1 | Jadwal penelitian . . . . .  | 20 |

## DAFTAR GAMBAR

|     |  |    |
|-----|--|----|
| .1  | Representasi dari MLP dengan dua <i>hidden layer</i> . . . . .                           | 8  |
| .2  | Arsitektur LeNet-5 adalah ANN bertipe CNN. . . . .                                       | 9  |
| .3  | Struktur sebuah <i>autoencoder</i> dengan tiga <i>fully-connected hidden layer</i> . . . | 9  |
| .4  | RBM dengan 5 <i>visible unit</i> dan 4 <i>hidden unit</i> . . . . .                      | 10 |
| .5  | DBN dengan satu <i>visible layer</i> dan tiga <i>hidden layer</i> . . . . .              | 11 |
| .6  | Ilustrasi sebuah CDBN. . . . .   | 12 |
| .1  | Skema tahapan penelitian. . . . .  | 13 |
| A.1 | Keluaran log CNN pada data MNIST . . . . .   | 26 |
| A.2 | Keluaran log CSAE pada data MNIST . . . . .  | 26 |
| A.3 | Grafik galat terhadap epoch CSAE pada MNIST . . . . .                                    | 27 |
| A.4 | Keluaran log CDBN pada data MNIST . . . . .  | 27 |

## DAFTAR LAMPIRAN

# PENDAHULUAN

## Latar Belakang

Inteligensi bisnis dan analitis (*Bussiness Intelligence and Analytics*, BIA) adalah tentang pengembangan teknologi, sistem, praktik, dan aplikasi untuk menganalisis data bisnis penting agar mendapatkan wawasan baru tentang bisnis dan pasar. Wawasan baru tersebut dapat digunakan untuk memperbaiki produk dan layanan, mencapai efisiensi operasional yang lebih baik, dan membina hubungan dengan pelanggan (Lim *et al.*, 2013). Dengan inisiatif BIA, bisnis mendapat wawasan dari pertumbuhan volume transaksi, produk, persediaan, pelanggan, pesaing, dan data industri yang dihasilkan oleh aplikasi skala perusahaan seperti, perencanaan sumber daya perusahaan (*Enterprise Resource Planning*, ERP), manajemen hubungan pelanggan (*Customer Relationship Management*, CRM), manajemen rantai pasok (*Supplay Chain Management*, SCM), manajemen pengetahuan (*knowledge management*), komputasi kolaboratif, analitik web, media sosial, dan data yang dibangkitkan oleh sensor (Chen *et al.*, 2012). Dua dari beberapa data yang dibangkitkan oleh sensor adalah data audio dan data video. Menurut konsep inteligensi bisnis terintegrasi yang diusulkan oleh Baars & Kemper (2008) bahwa wawasan dapat diambil dari penggabungan data terstruktur dan tak terstruktur.

Pada artikelnya, Helfenstein *et al.* (2014) memperkenalkan istilah *Emotional Business Intelligence* (EBI) yang bertujuan untuk mendukung bisnis yang terkait emosi dan keputusan yang peka terhadap emosi sebagai tambahan bagi pengambilan keputusan rasional. EBI memiliki misi membekali pelaku bisnis dengan wawasan dan sarana manajemen yang diperlukan. Menurut EBI berasal dari tiga domain utama: *emotional business*, *emotional intelligence* dan *business intelligence*. Untuk pemasar dan disainer dimungkinkan penggunaan teknik Kansei (*Kansei engineering*) lebih lanjut seperti yang terdapat pada Nagamachi & Lokman (2010).

Mehrabian (2008) menyatakan bahwa ekspresi wajah memberikan peran 55% terhadap keseluruhan pesan dari pembicaraanya, begitu juga komponen suara memberi peran 38%, sedangkan komponen verba menunjukkan peran hanya sekitar 7%. Hal ini memberikan indikasi kasar bahwa ekspresi wajah dan suara adalah dua hal yang paling penting pada transmisi emosi dari pembicaraanya. Di lain sisi, makna sistaksis dan semantik dari yang terucapkan tidak semua dapat tersampaikan. Penelitian klasik tertuju untuk menemukan jawaban atas pertanyaan “Apa yang dikatakan?” dan “Siapa yang mengatakan?” yang merupakan bahasan pada tema penelitian pengenalan ucapan (*speech recognition*) dan identifikasi pembicara (*speaker identification*). Tetapi baru-baru ini, pertanyaan “Bagaimana itu dikatakan?” telah menjadi perhatian lebih dari komunitas peneliti (Chen &

Luo, 2009). Picard (1995) mengistilahkannya dengan komputasi afektif (*affective computing*).

Pembelajaran dalam (*deep learning*) (Bengio *et al.* , 2009) merupakan cabang dari pembelajaran mesin (*machine learning*) berdasarkan satu set algoritma yang mencoba untuk membuat suatu model abstraksi pada tingkat tinggi terhadap data latih yang digunakan. Diantara arsitekturnya *deep neural networks*, *convolutional deep neural networks*, *deep belief networks*, dan *recurrent neural networks* telah diterapkan pada beberapa bidang seperti visi komputer (*computer vision*), pengenalan ucapan otomatis (*automatic speech recognition*), pengolahan bahasa alami (*natural language processing*), dan pengenalan audio (*audio recognition*) (LeCun *et al.* , 2015).

Pada penelitian ini kami akan menggunakan beberapa arsitektur pembelajaran dalam. Dan pada penelitian ini pula kami mengusulkan suatu model pembelajaran dalam agar dapat digunakan untuk pengenalan emosi berdasarkan modalitas ekspresi wajah, intonasi suara, dan ucapan tekstual yang menyertainya. Harapannya adalah dengan penelitian ini kami dapat membantu mengarahkan penelitian selanjutnya untuk memperluas kemampuan BIA serta menjembatani kesenjangan antara hasil penelitian yang telah ada sebelumnya dengan kendala yang ada di dunia bisnis dan industri.

## **Perumusan Masalah**

Dengan latar belakang dan acuan berfikir diatas, bahwa akan banyak data yang dapat dibangkitkan oleh sensor diantaranya data audio video. Dan salah satu manfaat jika data-data tersebut dapat dianalisis adalah dapat memberikan banyak pengertian dan pengetahuan terhadap emosi manusia yang terpancar saat berkomunikasi. Salah satu cabang dari pembelajaran mesin yang telah banyak diterapkan dengan baik di berbagai bidang pada masalah pengenalan pola (*pattern recognition*) adalah pembelajaran dalam. Sehingga pada penelitian ini kami mengusulkan suatu model pengenalan emosi berdasarkan modalitas ekspresi wajah, intonasi suara, dan ucapan tekstual yang menyertainya dengan menggunakan arsitektur pembelajaran dalam.

## **Tujuan Penelitian**

Tujuan dari penelitian ini adalah mengembangkan sebuah model pembelajaran dalam untuk pengenalan emosi berdasarkan pada ekspresi wajah, intonasi suara, dan ucapan tekstual.

## **Manfaat Penelitian**

Dengan penelitian ini kami berharap dapat memberikan sumbangan pengetahuan mengenai pengenalan emosi dari data audio video dan menjadi bagian dari analitis audio video pada inteligensi bisnis. Dengan penelitian ini pula kami mencoba membuka penelitian lanjutan mengenai perlombongan emosi (*emotion mining*) pada data audio video yang kaitannya untuk memperbaiki pemahaman kita terhadap emosi manusia dalam upaya meningkatkan kemampuan organisasi untuk mencapai tujuannya.

## **Ruang Lingkup Penelitian**

Penelitian ini membatasi ruang lingkupnya dengan menggunakan teknik pembelajaran dalam untuk mengenali emosi manusia berdasarkan ekspresi wajah, intonasi suara, dan ucapan tekstual dengan menggunakan dataset yang berisi kumpulan data audio video berbahasa inggris.

# TINJAUAN PUSTAKA

## Pengenalan Emosi

Beberapa penelitian terkait kasifikasi emosi dapat dibagi menjadi empat, yaitu: klasifikasi emosi dari wajah, klasifikasi emosi dari ucapan, klasifikasi emosi dari teks, dan klasifikasi emosi hibrida menggunakan gabungan diantara wajah, ucapan, dan teks.

### Pengenalan emosi dari suara

Artikel dari Meng *et al.* (2007) memperlihatkan pengklasifikasian emosi suara yang berkerja sebagai kombinasi dari sistem pengenalan suara otomatis. Algoritmanya menggunakan *Hidden Markov Models* (HMM) sebagai pengklasifikasi. Fitur yang dipertimbangkan untuk percobaannya terdiri dari 39 MFCC ditambah nada, intensitas dan tiga formant, termasuk beberapa turunan statistiknya.

Pada artikelnya Le & Provost (2013) menggunakan *Deep Belief Networks* (DBN) untuk memodelkan hubungan kompleks dan nonlinear antara fitur-fitur tingkat rendah. Ia telah mengusulkan sederetan pengklasifikasi hibrida terdiri dari *Hidden Markov Models* dan *Deep Belief Networks*. Karyanya menghasilkan bahwa parameter model yang optimal untuk emosi berbeda dengan parameter pada pengenalan suara dalam beberapa aspek penting. Dibandingkan dengan suara, pengenalan emosi diperlukan *learning rate* lebih kecil, lebih besar *input window* pada DBN, dan lebih sedikit HMM *states*. Pengamatannya menunjukkan bahwa emosi terletak pada ruang keputusan yang berbeda, jangka waktu *time window* lebih lama, dan kurang dipahami dinamika temporalnya dibandingkan dengan suara.

### Pengenalan emosi dari teks

Artikel dari Mishne *et al.* (2005) menghadirkan pekerjaan awal pada pengkalsifikasi teks blog sesuai dengan suasana hati penulisnya selama penulisan. Ia mempertimbangkan fitur tekstual berbeda, seperti jumlah frekuensi kata, polaritas emosional tulisan, panjang tulisan, *Point-wise Mutual Information* (PMI) memberikan bobot numerik untuk kata kunci berdasarkan hubungannya dengan suasana hati tertentu, penekanan kata, dan simbol-simbol khusus seperti emoticon dan tanda baca. Pengkalsifikasi yang digunakan dalam karyanya adalah *Support Vector Machine* (SVM).

Pada artikelnya Ezzat *et al.* (2012) menyajikan perlombongan teks (*text mining*) terhadap transkripsi rekaman audio untuk menemukan emosi pembicara. Dalam karyanya, ia

menggunakan metode seleksi fitur yang berbeda. Metode terbimbing dan tak terbimbing untuk kasifikasi teks. Bukti menunjukan keefektifan metode terbimbing untuk klasifikasi teks.

Pada artikelnya del Hoyo *et al.* (2009) menjelaskan tiga pendekatan yang berbeda: pendekatan statistik, pendekatan semantik, dan pendekatan statistik-semantik hibrida. Nilai *Term Frequency-Inverse Document Frequency* (TF-IDF) dari istilah dalam teks digunakan sebagai fitur dalam pendekatan statistik. Sistem statistik-semantik hibrida mencapai tingkat akurasi yang lebih besar dari pendekatan statistik ataupun semantik.

## Pengenalan emosi dari wajah

Pada artikelnya Liu *et al.* (2013) membangun arsitektur yang dalam, *Action Unit aware Deep Network* (AUDN), untuk pengenalan ekspresi wajah dengan memanfaatkan pengetahuan *prior* bahwa variasi kemunculan yang disebabkan oleh ekspresi dapat diurai menjadi sejumlah *Action Unit* (AU) lokal. AUDN yang diusulkan terdiri dari tiga modul secara berurutan, yaitu: Modul pertama, lapisan konvolusi dan lapisan *max-pooling*, yang bertujuan menghasilkan representasi lebih lengkap. Modul kedua lapisan reseptif AU dirancang untuk mencari subset pada representasi lengkapnya. Modul terakhir, *multi-layer Restricted Boltzmann Machines* (RBM) yang dimanfaatkan untuk mempelajari fitur hirarkis, yang kemudian menggabungkannya menjadi pengenalan ekspresi akhir.

Artikel dari Liu *et al.* (2014) menyajikan bentuk *Boosted Deep Belief Networks* (BDBN) pada tahap pelatihan. BDBN terdiri dari dua proses belajar yang berhubungan: proses *bottom-up unsupervised feature learning* (BU-UFL) yang belajar representasi fitur hirarkis pada data input yang diberikan dan proses *boosted top-down supervised feature strengthen* (BTD-SFS) yang memurnikan fitur-fitur gabungan tersebut dengan cara terbimbing.

## Pendekatan hibrida

Pada artikelnya Ngiam *et al.* (2011) menyajikan serangkaian tugas untuk pembelajaran multimodal dan menunjukkan bagaimana melatih *deep networks* yang belajar fitur untuk mengatasi tugas tersebut. Secara khusus, ia menunjukkan pembelajaran fitur modalitas bersilang, dimana fitur yang lebih baik untuk satu modalitas (misalnya video) dapat belajar jika beberapa modalitas (misalnya audio dan video) yang hadir pada saat pembelajaran fitur. Ia juga menunjukkan bagaimana cara belajar representasi bersama antara modalitas dan mengevaluasinya pada tugas yang unik, dimana pengkasifikasi dilatih hanya dengan data audio tapi diuji hanya dengan data video saja dan sebaliknya. Ia menggunakan data set CUAVE dan AVletters untuk memvalidasi klasifikasi ucapan audio visual.

Pada artikelnya Kim *et al.* (2013) menunjukkan efektifitas *Deep Belief Network* (DBN) untuk membangkitkan fitur emosi multimodal. DBN menangkap dependensi fitur non-linear dari fitur-fitur audio visual yang sekaligus mengurangi dimensi ruang fitur. Model generatifnya telah diterapkan pada mengolahan suara dan bahasa, demikian juga pada emosi. Dalam pengenalan emosi, Brueckner & Schuller (2012) menemukan bahwa penggunaan *Restricted Boltzmann Machine* (RBM) sebelum pada dua lapisan *nerual network* dengan *fine-tuning* dapat secara signifikan meningkatkan akurasi klasifikasi pada tantangan klasifikasi emosi Inter-speech (Schuller *et al.* , 2009). Karya oleh Stuhlsatz



*et al.* (2011) mengambil pendekatan berbeda untuk belajar fitur akustik dalam pengenalan emosi suara menggunakan *Generalize Discriminant Analysis* (GerDa) berdasarkan *Deep Neural Networks* (DNN).

Artikel dari Chuang & Wu (2004) menyajikan sebuah pendekatan untuk pengenalan emosi dari sinyal ucapan dan konten tekstual. Dalam analisis sinyal suaranya, beberapa fitur akustik diambil. Ia menggunakan *Principal Component Analysis* (PCA) untuk seleksi fitur sebagai representasi diskriminatif. *Support Vector Machines* (SVM) diadopsi untuk mengklasifikasikan kondisi emosional. Dalam analisis teks, semua kata kunci emosional dan kata-kata modifikasi didefinisikan secara manual. Tingkat intensitas emosi kata kunci emosional dan kata-kata modifikasi diperkirakan berdasarkan pada korpus emosi yang dikumpulkan. Keadaan emosional akhir ditentukan berdasarkan keluaran emosi dari analisis akustik dan tekstual.

Pada artikelnya Bhaskar *et al.* (2015) mengusulkan pendekatan untuk klasifikasi percakapan audio berdasarkan baik suara maupun teks. Kebaruan dalam pendekatan ini pilihan fitur dan pembangkitan sebuah vektor fitur tunggal untuk klasifikasi. Karya ini menggunakan metode standar seperti *Natural Language Processing* (NLP), WordNet Affect, SentiWordNet, dan *Support Vector Machines* (SVM). Ia menggunakan data set Semval-2007 dan eNTERFACE'05 EMOTION.

## **Representasi**

Representasi, fitur, atau ciri merupakan ....

### **Representasi pada citra**

Citra merupakan ....

### **Representasi pada suara**

Suara merupakan ...

### **Representasi pada teks**

Representasi teks sangat penting pada banyak penerapan, yang paling umum digunakan adalah secara *local representation*: *n-grams*, *bag-of-words*, *1-of-n coding*, dan secara *continuous representations*: *latent semantic analysis*, *latent dirichlet allocation*, *distributed representations*.

## ***Artificial Neural Networks***

*Artificial neural networks* (ANN) adalah sebuah model pembelajaran mesin (*machine learning*) yang terinspirasi oleh otak hewan (Hubel & Wiesel, 1968) dengan harapan ia dapat digunakan untuk kepentingan manusia dalam menyelesaikan beberapa permasalahan. Meskipun ANN terinspirasi dari jaringan neuron biologis, ANN pada kenyataannya

secara biologi jauh sekali. ANN merupakan bentuk yang sangat sederhana dari apa yang terjadi pada otak hewan (Jain *et al.* , 1996).

ANN pertama kali diperkenalkan pada tahun 1943 oleh Warren McCulloch and Walter Pitts, mereka bermaksud membuat model matematika dari sebuah otak (McCulloch & Pitts, 1943). Pada 1949 Donald Hebb memperkenalkan apa yang disebut sebagai *Hebbian learning*, biasanya ditujukan dengan sebuah slogan “Neuron that fire together wire together” (Hebb, 1949).

Sebuah ketertarikan meningkat pada bidang ini setelah Frank Rosenblatt memperkenalkan “*perceptron*” pada 1958. Ia menunjukkan bagaimana sebuah model matematika yang sederhana dapat digunakan untuk melatih dua lapisan jaringan (Rosenblatt, 1958). Namun, Seymour Papert dan Marvin Minsky membuka beberapa keterbatasan dari ANN, membuktikan bahwa fungsi eksklusif OR tidak bisa dipelajari oleh *perceptron*, bersamaan dengan itu mereka mendemonstrasikan bahwa tidak layak melatih ANN yang besar, dikarenakan keterbatasan sumber daya komputer saat itu (Minsky & Papert, 1969).

Setelah kekecewaan ini, pada penelitian ANN, ketertarikan meningkat kembali pada akhir tahun 70-an dan pada awal 80-an karena penemuan algoritma *backpropagation* yang dapat memecahkan masalah eksklusif OR dan karena telah majunya arsitektur komputer saat itu. Pada tahun 90-an metode pembelajaran mesin yang lain seperti *Support Vector Machines* (SVM) (Cortes & Vapnik, 1995) menjadi populer. SVM memiliki penjelasan teori yang jelas dan menyelesaikan masalah secara lebih baik dari ANN. Sejak tahun 2000-an, ketertarikan terhadap ANN kembali meningkat dikarenakan kemajuan pada *Convolutional Neural Networks* (CNN, atau LeNet setelah LeCun *et al.* (1998a) memperbaiki disain awal yang dibuat oleh Fukushima (1988)), *Unsupervised Feature Learning*, *Stacked AutoEncoder* dan *Deep Belief Networks* yang diusulkan oleh Geoffrey Hinton di tahun 2006 (Hinton *et al.* , 2006). Berikut merupakan beberapa tipe ANN terkait penelitian ini yang akan digunakan sebagai pengklasifikasi.

## ***Convolutional Neural Networks***

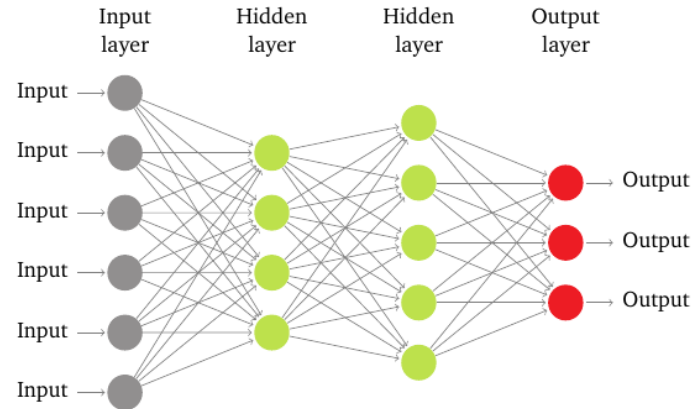
### ***Multilayer Perceptrons (MLP)***

*Multilayer Perceptrons* (MLP), salah satu tipe ANN, juga sering disebut *feedforward neural networks* atau *deep feedforward networks*, adalah model pembelajaran dalam klasik. Tujuan dari MLP adalah untuk memperkirakan fungsi  $f^*$ . Sebagai contoh, untuk sebuah pengklasifikasi,  $y = f^*(x)$  memetakan sebuah input vektor  $x$  terhadap sebuah kategori  $y$ . Sebuah MLP mendefinisikan sebuah fungsi pemetaan  $y = f(x, \theta)$  dan belajar dari nilai parameter  $\theta$  dari hasil pada aproksimasi terbaik fungsi tersebut (Goodfellow *et al.* , 2016).

MLP dengan *hidden layer* menurut teorema *universal approximation* (Hornik *et al.* , 1989) menyediakan sebuah kerangka kerja hampiran universal. Teorema tersebut menyatakan MLP dengan sebuah lapisan keluaran linear dan setidaknya satu *hidden layer* dengan fungsi aktivasi skuas (*squashing*, seperti fungsi aktivasi logistik sigmoid) dapat menghampiri setiap fungsi *Borel measurable* dari satu dimensi ruang terhingga kepada yang lain dengan sejumlah galat yang diharapkan, jika disediakan jaringan *hidden units* yang cukup.

Struktur dari MLP dapat dilihat sebagai sebuah graf. Pada konteks ini simpulnya

disebut simpul atau neuron dan tepiannya disebut koneksi sinapsis. Karakteristik penting MLP bahwa setiap koneksi sinapsis memiliki kekuatan masing-masing (disebut bobot) yang disesuaikan saat proses pelatihan.



Gambar .1: Representasi dari MLP dengan dua *hidden layer*.

Seperti terlihat pada Gambar 1, jaringannya terstruktur pada lapisan-lapisan (*layers*), dengan koneksi antara simpulnya. Secara umum simpul didalam lapisan yang sama tidak terhubung dengan simpul lainnya pada lapisan yang sama, tetapi terhubung pada simpul-simpul pada lapisan setelahnya.

Lapisan-lapisan tersebut dapat dikelompokkan menjadi, *input layer* yang menerima masukan data yang diberikan, *hidden layer* yang mendefinisikan sebuah representasi (fitur) dari data yang diberikan oleh *input layer*, dan *output layer* merupakan keluaran setelah dilewatkan pada jaringan.

Data masukan ditransformasikan kedalam sebuah vektor bilangan ril dan disampaikan pada jaringan. Setelah data dilewatkan maka neuron menjadi teraktifasi. Secara matematika kondisi ini dapat dimodelkan sebagai berikut:

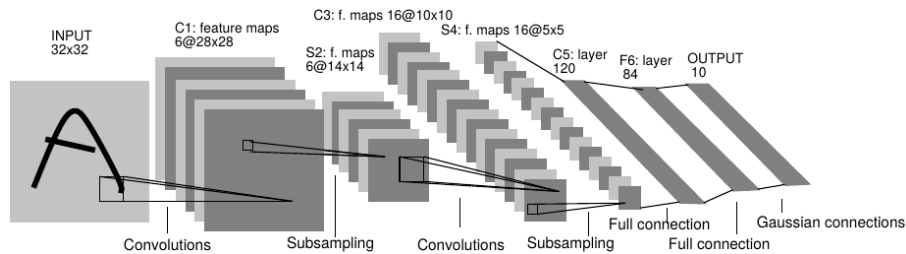
$$y_i^{(l+1)} = \sigma \left( \sum_j w_{ij} y_j^{(l)} + b_i \right) \quad (.1)$$

Disini,  $b_1$  adalah bias yang diasosiasikan dengan unit dan  $\sigma$  mengacu pada fungsi aktifasi. Fungsi aktifasi memiliki peran yang penting pada jaringan, ia digunakan untuk membatasi kisaran nilai yang dihasilkan. Biasanya fungsi aktifasi yang digunakan dari keluarga sigmoid (seperti logistic sigmoid dan tanh).

## ***Convolutional Neural Networks (CNN)***

*Convolutional Neural Networks* (CNN, ConvNet) adalah sebuah ANN dan merupakan varian dari MLP yang didisain untuk mengenali pola visual langsung dari piksel citra dengan meminimalkan praproses (LeCun *et al.* , 1998a). Nama *convolutional* mengindikasikan bahwa jaringan tersebut menggunakan operasi matematika yang disebut *convolution*, selain itu CNN juga menggunakan operasi yang disebut *pooling* (*subsampling*).

Sehingga pada dasarnya CNN terdiri dari satu atau beberapa *convolutional layer* dan *pooling layer* yang ditumpuk lalu diteruskan dengan satu atau beberapa *fully-connected layer* seperti pada MLP. Gambar 2 merupakan contoh sebuah arsitektur CNN.

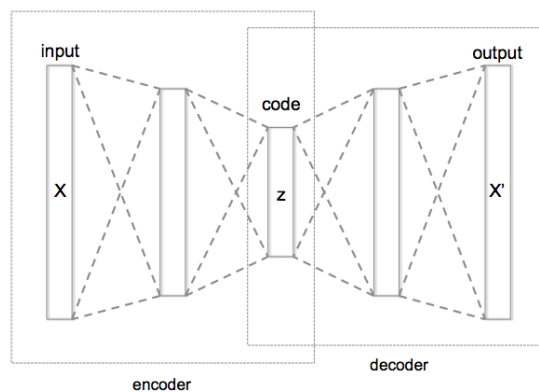


Gambar .2: Arsitektur LeNet-5 adalah ANN bertipe CNN.

## Stacked Convolutional AutoEncoder

### AutoEncoder (AE)

*Autoencoder* (AE) (Bengio *et al.* , 2009) adalah sebuah ANN yang digunakan untuk pembelajaran tak terbimbing secara efisien (Liou *et al.* , 2014). Tujuan dari *autoencoder* adalah untuk belajar sebuah *representation* (*encoding*) dari sehimpunan data. Secara aritektur, bentuk sederhana dari sebuah *autoencoder* adalah sebuah *feedforward, non-recurrent neural network* sangat mirip dengan MLP, memiliki satu *input layer*, satu *output layer*, tetapi dengan jumlah simpul *output layer* sama seperti pada *input layer*-nya, dengan maksud merekonstruksi inputnya sendiri (daripada memprediksi nilai target  $Y$  dengan input  $X$ ) dan selalu terdiri dari dua bagian, bagian *encoder* dan decoder.



Gambar .3: Sturktur sebuah *autoencoder* dengan tiga *fully-connected hidden layer*.

### Stacked AutoEncoder (SAE)

*Stacked AutoEncoder* (SAE) adalah ...

## ***Stacked Convolutional AutoEncoder (SCAE)***

*Stacked Convolutional AutoEncoder (SCAE)* adalah ...

## ***Convolutional Deep Belief Networks***

### ***Restricted Boltzman Machine (RBM)***

*Restricted Boltzmann Machine (RBM)* (Smolensky, 1986) adalah ANN yang bersifat stokastik generatif yang dapat belajar dari sebuah distribusi probabilitas melalui sejumlah set inputan. RBM dapat dilatih dengan cara terbimbing ataupun tidak, bergantung pada kasusnya. RBM merupakan varian dari algoritma *Boltzmann machines*, dengan dibatasi (*restricted*) bahwa pada setiap layernya tidak ada yang terhubung satu sama lainnya. Beberapa aplikasi dari RBM diantaranya pada *dimensional reduction* (Hinton & Salakhutdinov, 2006), *classification* (Larochelle & Bengio, 2008), *collaborative filtering* (Salakhutdinov *et al.* , 2007), *feature learning* (Coates *et al.* , 2010), dan *topic modeling* (Hinton & Salakhutdinov, 2009).

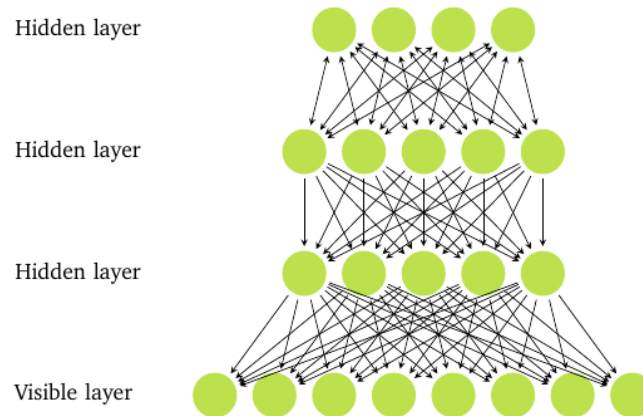


Gambar .4: RBM dengan 5 *visible unit* dan 4 *hidden unit*.

### ***Deep Belief Networks (DBN)***

*Deep Belief Network (DBN)* atau *Stacked Restricted Boltzmann Machine (SRBM)* adalah sebuah tipe ANN yang bersifat generatif probabilistik yang tersusun dari beberapa *layer hidden unit*, dengan adanya hubungan antara *layer*-nya tetapi tidak antar *unit* dalam satu *layer*. Jika dilatih pada sejumlah data secara tak terbimbing, DBN dapat merekonstruksi secara probabilistik *input*-nya. *Layer*-nya bekerja sebagai detektor fitur pada masukannya (Hinton, 2009). Setelah tahap pembelajaran ini, DBN selanjutnya dapat dilatih secara terbimbing untuk melakukan klasifikasi (Hinton *et al.* , 2006).

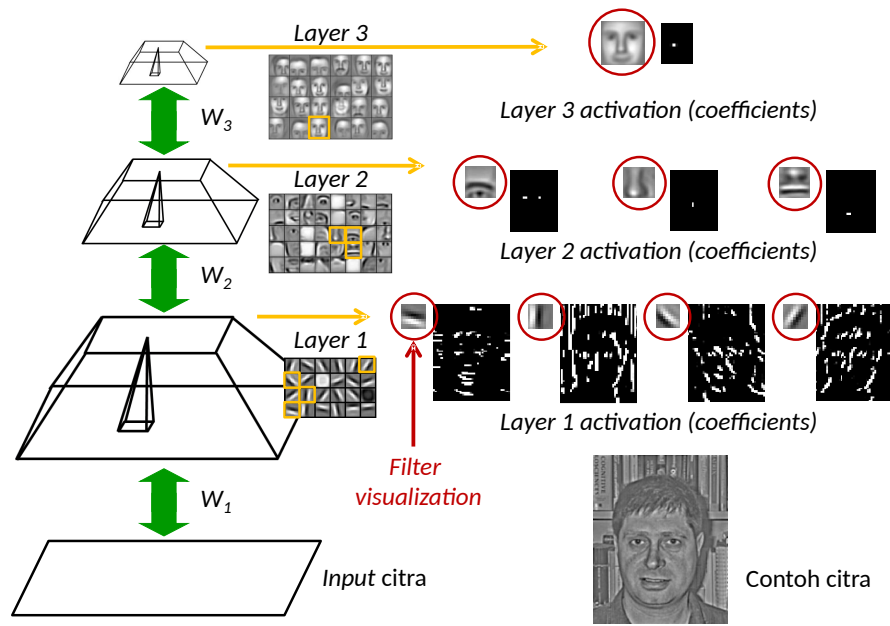
DBN dapat dilihat sebagai komposisi sederhana, jaringan tak terbimbing seperti RBM atau *autoencoder* (Bengio *et al.* , 2007), dimana setiap *sub-network* pada *hidden layer* berlaku sebagai *visible layer* untuk *layer* berikutnya. Ini juga mengarah pada, prosedur pelatihan tak terbimbing *layer* demi *layer*, dimana diterapkan *contrastive divergence* (Hinton, 2002) untuk setiap *sub-network* pada gilirannya, dimulai dari pasangan *layer* terendah. Pada pengamatannya Yee-Whye Teh (Hinton *et al.* , 2006), bahwa DBN dapat dilatih secara *greedy*, satu *layer* pada satu waktu, membuatnya sebagai algoritma pembelajaran dalam pertama yang efektif (Bengio *et al.* , 2009). Gambar 4 berikut memperlihatkan arsitekur sebuah DBN.



Gambar .5: DBN dengan satu *visible layer* dan tiga *hidden layer*.

### ***Convolutional Deep Belief Networks (CDBN)***

*Convolutional Deep Belief Networks* (CDBN) adalah sebuah tipe ANN yang tersusun dari beberapa lapisan *Convolutional Restricted Boltzmann Machines* (CRBM) yang ditumpuk bersama (Lee *et al.* , 2009a). CDBN memiliki struktur yang sangat mirip dengan *Convolutional Neural Networks* (CNN) dan dilatih mirip dengan DBN. Sehingga CDBN menggunakan struktur citra 2D, seperti yang dilakukan CNN, dan menggunakan *pre-training* seperti DBN. CDBN menyediakan struktur generik yang dapat digunakan secara efektif pada tugas pengolahan citra dan pengenalan objek (Lee *et al.* , 2009b). CDBN menggunakan teknik *probabilistic max-pooling* untuk mereduksi dimensi pada *layer* yang lebih tinggi pada jaringan. Pelatihan pada jaringannya melibatkan tahap *pre-training* yang diselesaikan dengan cara *greedy layer-wise* seperti pada DBN. Bergantung pada, apakah jaringannya digunakan untuk tugas diskriminasi atau generatif. Proses selanjutnya adalah *fine tuned* atau *training* baik dengan algoritma *back-propagation* atau *up-down*.

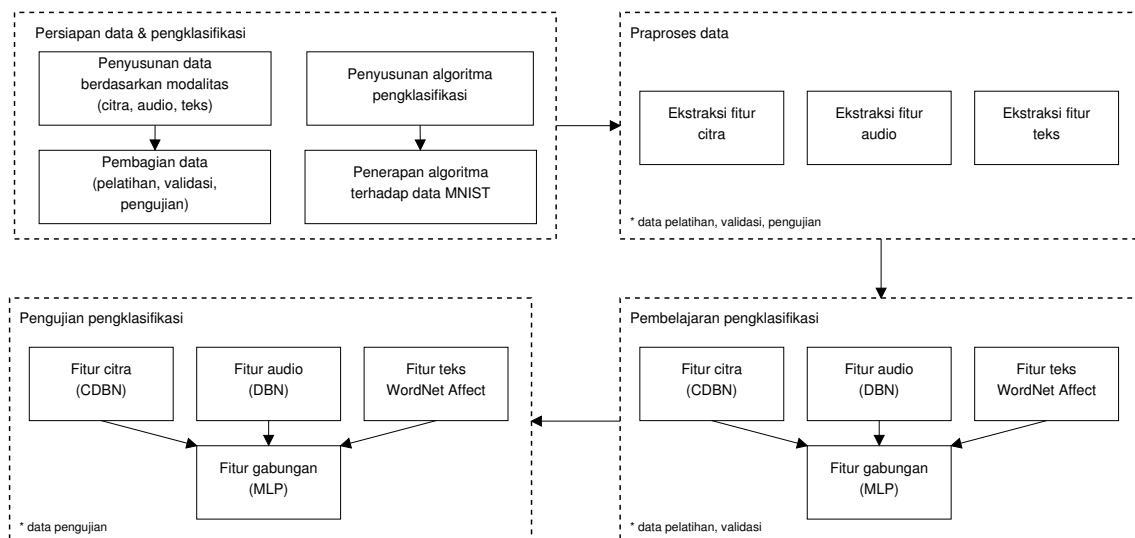


Gambar .6: Ilustrasi sebuah CDBN.

# METODE

## Tahap Penelitian

Penelitian ini akan terdiri dari tiga tahapan, terlihat pada Gambar 6, yaitu: 1) persiapan, 2) praproses, 3) proses pembelajaran, 4) proses pengujian. Tahap persiapan terdiri dari penyusunan data modalitas yang terdiri dari tiga jenis data, video, audio, dan teks. Masing-masing jenis data tersebut dibagi menjadi tiga bagian, sebagai data pelatihan, validasi, dan pengujian. Pada tahap ini juga disusun algoritma pengklasifikasi dan kemudian diterapkan untuk pengenalan tulisan tangan pada data set MNIST (LeCun *et al.* , 1998b). Tahap selanjutnya adalah praproses. Pada tahap ini dilakukan ekstraksi fitur pada masing-masing data tersebut. Selanjutnya yaitu tahap pembelajaran. Di tahap ini data tersebut kemudian dikenakan proses pembelajaran menggunakan beberapa pengklasifikasi sebagai pembanding, seperti CDBN, CSAE, atau CNN dan dengan WordNet Affect (WNA) (Strapparava *et al.* , 2004) pada data teks. Keluaran pada setiap pengklasifikasi tersebut kemudian dijadikan sebagai fitur masukan untuk pengklasifikasi fitur gabungan dengan fungsi Softmax. Dan tahap terakhir adalah tahap pengujian pengklasifikasi tersebut terhadap data pengujian.



Gambar .1: Skema tahapan penelitian.



## Alat dan Data

Banyak data set yang tersedia secara daring dan untuk karya ini akan digunakan dua diantaranya. Pertama adalah sebuah data audio video yang diberi nama eNTERFACE'05 EMOTION<sup>1</sup> (Martin *et al.* , 2006) sedangkan yang kedua adalah *Interactive Emotional Dyadic Motion Capture* (IEMOCAP)<sup>2</sup> (Busso *et al.* , 2008). Pada kedua data set tersebut, aktor melakukan adegan improvisasi emosional dari skenario dan naskah yang diberikan. Keduanya dapat digunakan sebagai data set referensi untuk pengujian dan evaluasi algoritma pengenalan emosi pada video, audio, ataupun audio video. Adapun alat yang digunakan adalah sebuah komputer laptop dengan prosesor Intel Core i3 dan RAM 8 Gb yang bersistem operasi Linux Fedora 24. Sedangkan perangkat lunak yang digunakan adalah Lyx (2.2) *Document processor*, KBibTex (0.6) *Bibtex editor*, LibreOffice (5.1) *Office suite*, IPython Notebook (4.2) *Interactive python web application* dan Spyder IDE (3.1) *Scientific python development environment*. Dalam implementasinya, penelitian ini menggunakan Python (2.7)<sup>3</sup> sebagai bahasa pemrogramannya dan Neupy (0.5)<sup>4</sup> sebagai kerangka kerjanya (*framework*) dan beberapa pustaka seperti Theano (0.8)<sup>5</sup>, Scikit-learn (0.18)<sup>6</sup>, Numpy (1.1)<sup>7</sup>, Matplotlib (1.5)<sup>8</sup>, NLTK (3.2)<sup>9</sup>, Praat (5.3)<sup>10</sup>, dan OpenCV (2.4)<sup>11</sup>.

## Persiapan

Tahap ini terdiri dari proses penyusunan data dan penyusunan implementasi algoritma pengklasifikasi. Penyusunan data dilakukan dengan mengekstrak data menjadi tiga tipe yaitu, data citra yang diambil dari data video menjadi beberapa bagian citra yang memiliki ekspresi wajah, data suara yang diambil dari data audio yang memiliki intonasi, dan data teks yang diambil dari naskah skenarionya. Selanjutnya ketiga tipe data tersebut dibagi menjadi tiga bagian sesuai fungsinya yaitu, untuk pelatihan 50%, validasi 20%, dan pengujian 30%. Masing-masing data tersebut kemudian diberi *label* emosinya dapat berupa *anger*, *disgust*, *fear*, *happiness*, *sadness*, ataupun *surprise*.

Penyusunan implementasi algoritma pengklasifikasi pada karya ini menggunakan bahasa pemrograman Python versi 2.7, alasannya karena banyaknya komunitas peneliti pembelajaran mesin (*machine learning*) menggunakannya serta tersedianya pustaka pendukung sehingga memudahkan proses penyusunan, dan tentunya ia menggunakan prinsip kode terbuka (*open source*) yang sangat mendukung kemajuan ilmu pengetahuan dan teknologi. Sebelum pengklasifikasi digunakan dilakukan proses percobaan terlebih dahulu pada data MNIST, untuk pengklasifikasian citra tulisan tangan numerik dari 0 sampai 9, sebagai praevaluasi bagi algoritma yang digunakan.

---

<sup>1</sup><http://enterface.net/enterface05/>

<sup>2</sup><http://sail.usc.edu/iemocap/>

<sup>3</sup><http://www.python.org/>

<sup>4</sup><http://neupy.com/>

<sup>5</sup><http://deeplearning.net/software/theano/>

<sup>6</sup><http://scikit-learn.org/>

<sup>7</sup><http://numpy.org/>

<sup>8</sup><http://matplotlib.org/>

<sup>9</sup><http://nltk.org/>

<sup>10</sup><http://fon.hum.uva.nl/praat/>

<sup>11</sup><http://opencv.org/>

## Praproses

Tahap praproses sendiri merupakan proses ekstraksi fitur atau representasi terhadap data citra, suara, dan teks. Proses ekstraksi fitur data citra dilakukan dengan konversi citra menjadi citra abu-abu (*gray scaling*), kemudian pemotongan citra wajah (*cropping*) dengan deteksi wajah Haar Cascades (Viola & Jones, 2001), selanjutnya dilakukan *whitening* dan normalisasi sehingga terbentuk sebuah matrik fitur citra, proses tersebut dilakukan dengan bantuan pustaka OpenCV. Proses ekstraksi fitur data suara dilakukan dengan transformasi sinyal audio menjadi bentuk spektrogram dan MFCC (...) dan kemudian dilakukan *whitening*, karena berada dalam domain frekuensi dan domain cepstral terhadap waktu sehingga ia dapat direpresentasikan menjadi sebuah matrik fitur suara, proses tersebut dilakukan dengan bantuan pustaka Praat. Untuk data teks digunakan pendekatan *word embedding* sehingga setiap kata akan direpresentasikan sebagai sebuah vektor dengan pemodelan seperti word2vec (Mikolov et al., 2013). Kata merupakan pembentuk kalimat, dalam hal ini teks tersebut merupakan sebuah kalimat yang kemudian direpresentasikan sebagai sebuah matrik fitur teks. Proses tersebut dapat dilakukan dengan bantuan pustaka NLTK. Ketiga matrik fitur tersebut kemudian dapat dikenakan proses pembelajaran dan pengujian baik secara terpisah maupun kombinasi gabungannya.

## Pembelajaran

Tahap pembelajaran merupakan tahap inti dari karya ini. Secara sederhana tahap ini merupakan proses penghampiran suatu fungsi non linear terhadap data fitur yang diberikan (*universal approximation*), sehingga fungsi tersebut dapat digunakan sebagai suatu fungsi pembeda (*discrimination function*) yang dapat mengelompokkan data menjadi beberapa gugus atau kelas. Proses yang terjadi pada tahap ini adalah pembelajaran terhadap matrik fitur dari data citra, suara, dan teks dengan menggunakan pengklasifikasi ANN tipe CDBN dengan dua atau tiga hidden layer serta tipe lainnya seperti SCAE dan CNN sebagai pembanding.

### Arsitektur CNN

Berikut ini merupakan arsitektur jaringan *Convolutional Neural Networks* (CNN) yang akan digunakan dengan menggunakan algoritma pembelajaran AdaDelta (Zeiler, 2012) yang merupakan sebuah varian dari algoritma *Stochastic Gradient Descent* (SGD) yang menggunakan *adaptive learning rate*.

| No | Masukan      | Tipe Layer         | Keluaran     |
|----|--------------|--------------------|--------------|
| 1  | (1, 28, 28)  | <i>Input</i>       | (1, 28, 28)  |
| 2  | (1, 28, 28)  | <i>Convolution</i> | (32, 26, 26) |
| 3  | (32, 26, 26) | <i>Batch Norm</i>  | (32, 26, 26) |
| 4  | (32, 26, 26) | <i>Relu</i>        | (32, 26, 26) |
| 5  | (32, 26, 26) | <i>Convolution</i> | (48, 24, 24) |
| 6  | (48, 24, 24) | <i>Batch Norm</i>  | (48, 24, 24) |
| 7  | (48, 24, 24) | <i>Relu</i>        | (48, 24, 24) |
| 8  | (48, 24, 24) | <i>Max Pooling</i> | (48, 12, 12) |
| 9  | (48, 12, 12) | <i>Convolution</i> | (64, 10, 10) |
| 10 | (64, 10, 10) | <i>Batch Norm</i>  | (64, 10, 10) |
| 11 | (64, 10, 10) | <i>Relu</i>        | (64, 10, 10) |
| 12 | (64, 10, 10) | <i>Max Pooling</i> | (64, 5, 5)   |
| 13 | (64, 5, 5)   | <i>Reshape</i>     | 1600         |
| 14 | 1600         | <i>Linear</i>      | 1024         |
| 15 | 1024         | <i>Batch Norm</i>  | 1024         |
| 16 | 1024         | <i>Relu</i>        | 1024         |
| 17 | 1024         | <i>Softmax</i>     | 10           |

Tabel .1: Arsitektur CNN beserta tipe setiap *layer*-nya (cont. MNIST, 28x28 piksel)

Secara sederhana *layer* pada arsitektur CNN ini terdiri dari *Convolution*, *Convolution*, *Subsampling*, *Convolution*, *Subsampling*, *Full connection*, dan terakhir dengan fungsi *Softmax*. Sebagai uraian lebih jelas pada masing-masing tipe *layer* atau proses dan algoritma tersebut dapat dilihat pada laman lampiran.

## Arsitektur SCAE

Berikut ini merupakan arsitektur jaringan *Stacked (Deep) Convolutional AutoEncoder* (SCAE) yang akan digunakan menggunakan algoritma pembelajaran ...

| No | Masukan | Tipe <i>Layer</i> | Keluaran |
|----|---------|-------------------|----------|
| 1  |         |                   |          |
| 2  |         |                   |          |
| 3  |         |                   |          |
| 4  |         |                   |          |
| 5  |         |                   |          |
| 6  |         |                   |          |
| 7  |         |                   |          |
| 8  |         |                   |          |
| 9  |         |                   |          |
| 10 |         |                   |          |
| 11 |         |                   |          |
| 12 |         |                   |          |
| 13 |         |                   |          |
| 14 |         |                   |          |
| 15 |         |                   |          |
| 16 |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |
|    |         |                   |          |

Tabel .2:

## Arsitektur CDBN

Arsitektur jaringan *Convotutional Deep Belief Netoworks* (CDBN) yang akan digunakan adalah sebagai berikut:

## Pengujian

Setelah melalui proses pembelajaran pada tahap sebelumnya pengklasifikasi, dengan nilai bobot dan bias optimal pada setiap *layer*-nya, diterapkan pada matrik fitur data uji. Pada dasarnya proses ini sama seperti pada proses pembelajaran, menggunakan algoritma CDBN, SCAE, ataupun CNN, yang membedakan adalah matrik fitur yang digunakan berasal dari data uji citra, suara, dan teks. Tahap ini bertujuan untuk mengetahui seberapa efektif dan akurat pengklasifikasi dalam mengklasifikasikan data uji sesuai dengan *label* emosinya sekaligus membuktikan secara empiris kemampuan ketiga tipe *convolutional deep learning* ini dalam mengklasifikasikan emosi berdasarkan tiga modalitas tersebut.

Setelah melalui proses pengujian dengan data uji, proses terakhir dari tahap ini adalah *live demo* yang diambil dari kamera web (*webcam*) secara *live* dengan menggunakan

sekenario yang sama dengan data set yang digunakan. Proses ini dilakukan untuk mengetahui secara efektif apakah dengan data set yang telah digunakan sudah cukup mewakili jika diterapkan di lapangan dan sebagai masukan untuk penelitian selanjutnya.

# HASIL DAN PEMBAHASAN

## MNIST

Berikut ini merupakan tabel ...

| Pengklasifikasi | MNIST  | Face |
|-----------------|--------|------|
| MLP             | 0.9832 |      |
| CNN             | 0.9898 |      |
| SAE             |        |      |
| SCAE            | 0.9233 |      |
| DBN             |        |      |
| CDBN            |        |      |
| RNN             |        |      |

Tabel .1: Tabel akurasi pengklasifikasi terhadap data set

## Pembahasan

Berdasarkan hasil sementara tersebut ...

# RENCANA DAN PERKEMBANGAN

Berikut ini merupakan tabel jadwal penelitian yang direncanakan.

| No | Aktifitas Penelitian         | Tahun/Bulan |     |     |     |      |     |     |     |     |     |
|----|------------------------------|-------------|-----|-----|-----|------|-----|-----|-----|-----|-----|
|    |                              | 2016        |     |     |     | 2017 |     |     |     |     |     |
|    |                              | Sep         | Okt | Nov | Des | Jan  | Feb | Mar | Apr | Mei | Jun |
| 1  | Studi literatur              | x           | x   | x   | x   | x    |     |     |     |     |     |
| 2  | Sidang komisi I              |             |     |     |     |      |     | x   |     |     |     |
| 3  | Pendaftaran kolokium         |             |     |     |     |      |     | x   |     |     |     |
| 4  | Kolokium                     |             |     |     |     |      |     | x   |     |     |     |
| 5  | Pengembangan dan pengujian   |             |     |     |     |      | x   | x   | x   |     |     |
| 6  | Evaluasi dan analisis        |             |     |     |     |      |     |     | x   |     |     |
| 7  | Publikasi jurnal             |             |     |     |     |      |     |     | x   |     |     |
| 8  | Tesis dan registrasi seminar |             |     |     |     |      |     |     | x   |     |     |
| 9  | Sidang komisi II             |             |     |     |     |      |     |     |     | x   |     |
| 10 | Seminar                      |             |     |     |     |      |     |     |     | x   |     |
| 11 | Sidang komisi III            |             |     |     |     |      |     |     |     |     | x   |
| 12 | Ujian tertutup               |             |     |     |     |      |     |     |     |     | x   |

Tabel .1: Jadwal penelitian

# Bibliografi

- Baars, Henning, & Kemper, Hans-George. 2008. Management support with structured and unstructured data—an integrated business intelligence framework. *Information Systems Management*, **25**(2), 132–148. (document)
- Bengio, Yoshua, Lamblin, Pascal, Popovici, Dan, Larochelle, Hugo, *et al.* . 2007. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, **19**, 153. (document)
- Bengio, Yoshua, *et al.* . 2009. Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, **2**(1), 1–127. (document)
- Bhaskar, Jasmine, Sruthi, K, & Nedungadi, Prema. 2015. Hybrid approach for emotion classification of audio conversation based on text and speech mining. *Procedia Computer Science*, **46**, 635–643. (document)
- Brueckner, Raymond, & Schuller, Björn. 2012. Likability Classification-A Not so Deep Neural Network Approach. *Page 290–293 of: INTERSPEECH*. Citeseer. (document)
- Busso, Carlos, Bulut, Murtaza, Lee, Chi-Chun, Kazemzadeh, Abe, Mower, Emily, Kim, Samuel, Chang, Jeannette N, Lee, Sungbok, & Narayanan, Shrikanth S. 2008. IEMO-CAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, **42**(4), 335–359. (document)
- Chen, Hsinchun, Chiang, Roger HL, & Storey, Veda C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, **36**(4), 1165–1188. (document)
- Chen, Shi-Huang, & Luo, Yu-Ren. 2009. Speaker verification using MFCC and support vector machine. *Page 18–20 of: Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1. (document)
- Chuang, Ze-Jing, & Wu, Chung-Hsien. 2004. Emotion recognition using acoustic features and textual content. *Page 53–56 of: Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, vol. 1. IEEE. (document)
- Coates, Adam, Lee, Honglak, & Ng, Andrew Y. 2010. An analysis of single-layer networks in unsupervised feature learning. *Ann Arbor*, **1001**(48109), 2. (document)
- Cortes, Corinna, & Vapnik, Vladimir. 1995. Support-vector networks. *Machine learning*, **20**(3), 273–297. (document)



- del Hoyo, Rafael, Hupont, Isabelle, Lacueva, Francisco J, & Abadía, David. 2009. Hybrid text affect sensing system for emotional language analysis. *Page 3 of: Proceedings of the international workshop on affective-aware virtual agents and social robots*. ACM. (document)
- Ezzat, Souraya, El Gayar, Neamat, & Ghanem, M. 2012. Sentiment analysis of call centre audio conversations using text classification. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl*, **4**(1), 619–627. (document)
- Fukushima, Kunihiro. 1988. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, **1**(2), 119–130. (document)
- Goodfellow, Ian, Bengio, Yoshua, & Courville, Aaron. 2016. *Deep learning*. MIT Press. (document)
- Hebb, Donald O. 1949. *The organisation of behaviour*. (document)
- Helfenstein, Sacha, Kaikova, Olena, Khriyenko, Oleksiy, & Terziyan, Vagan. 2014. Emotional Business Intelligence. *Page 14–21 of: Human System Interactions (HSI), 2014 7th International Conference on*. IEEE. (document)
- Hinton, Geoffrey E. 2002. Training products of experts by minimizing contrastive divergence. *Neural computation*, **14**(8), 1771–1800. (document)
- Hinton, Geoffrey E. 2009. Deep belief networks. *Scholarpedia*, **4**(5), 5947. (document)
- Hinton, Geoffrey E, & Salakhutdinov, Ruslan R. 2006. Reducing the dimensionality of data with neural networks. *science*, **313**(5786), 504–507. (document)
- Hinton, Geoffrey E, & Salakhutdinov, Ruslan R. 2009. Replicated softmax: an undirected topic model. *Page 1607–1614 of: Advances in neural information processing systems*. (document)
- Hinton, Geoffrey E, Osindero, Simon, & Teh, Yee-Whye. 2006. A fast learning algorithm for deep belief nets. *Neural computation*, **18**(7), 1527–1554. (document)
- Hornik, Kurt, Stinchcombe, Maxwell, & White, Halbert. 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, **2**(5), 359–366. (document)
- Hubel, David H, & Wiesel, Torsten N. 1968. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, **195**(1), 215–243. (document)
- Jain, Anil K, Mao, Jianchang, & Mohiuddin, K Moidin. 1996. Artificial neural networks: A tutorial. *Computer*, **29**(3), 31–44. (document)
- Kim, Yelin, Lee, Honglak, & Provost, Emily Mower. 2013. Deep learning for robust feature generation in audiovisual emotion recognition. *Page 3687–3691 of: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. (document)

- Larochelle, Hugo, & Bengio, Yoshua. 2008. Classification using discriminative restricted Boltzmann machines. *Page 536–543 of: Proceedings of the 25th international conference on Machine learning*. ACM. (document)
- Le, Duc, & Provost, Emily Mower. 2013. Emotion recognition from spontaneous speech using hidden markov models with deep belief networks. *Page 216–221 of: Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*. IEEE. (document)
- LeCun, Yann, Bottou, Léon, Bengio, Yoshua, & Haffner, Patrick. 1998a. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11), 2278–2324. (document)
- LeCun, Yann, Cortes, Corinna, & Burges, Christopher JC. 1998b. *The MNIST database of handwritten digits*. (document)
- LeCun, Yann, Bengio, Yoshua, & Hinton, Geoffrey. 2015. Deep learning. *Nature*, **521**(7553), 436–444. (document)
- Lee, Honglak, Grosse, Roger, Ranganath, Rajesh, & Ng, Andrew Y. 2009a. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Page 609–616 of: Proceedings of the 26th annual international conference on machine learning*. ACM. (document)
- Lee, Honglak, Pham, Peter, Largman, Yan, & Ng, Andrew Y. 2009b. Unsupervised feature learning for audio classification using convolutional deep belief networks. *Page 1096–1104 of: Advances in neural information processing systems*. (document)
- Lim, Ee-Peng, Chen, Hsinchun, & Chen, Guoqing. 2013. Business intelligence and analytics: Research directions. *ACM Transactions on Management Information Systems (TMIS)*, **3**(4), 17. (document)
- Liou, Cheng-Yuan, Cheng, Wei-Chen, Liou, Jiun-Wei, & Liou, Daw-Ran. 2014. Autoencoder for words. *Neurocomputing*, **139**, 84–96. (document)
- Liu, Mengyi, Li, Shaoxin, Shan, Shiguang, & Chen, Xilin. 2013. Au-aware deep networks for facial expression recognition. *Page 1–6 of: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE. (document)
- Liu, Ping, Han, Shizhong, Meng, Zibo, & Tong, Yan. 2014. Facial expression recognition via a boosted deep belief network. *Page 1805–1812 of: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (document)
- Martin, Olivier, Kotsia, Irene, Macq, Benoit, & Pitas, Ioannis. 2006. The eNTERFACE'05 audio-visual emotion database. *Page 8–8 of: 22nd International Conference on Data Engineering Workshops (ICDEW'06)*. IEEE. (document)
- McCulloch, Warren S, & Pitts, Walter. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, **5**(4), 115–133. (document)

- Mehrabian, Albert. 2008. Communication without words. *Communication Theory*, 193–200. (document)
- Meng, Hong, Pittermann, Johannes, Pittermann, Angela, & Minker, Wolfgang. 2007. Combined speech-emotion recognition for spoken human-computer interfaces. *Page 1179–1182 of: Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on*. IEEE. (document)
- Mikolov, Tomas, Chen, Kai, Corrado, Greg, & Dean, Jeffrey. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. (document)
- Minsky, Marvin, & Papert, Seymour. 1969. Perceptrons. (document)
- Mishne, Gilad, *et al.* . 2005. Experiments with mood classification in blog posts. *Page 321–327 of: Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access*, vol. 19. Citeseer. (document)
- Nagamachi, M, & Lokman, AM. 2010. *Innovations of Kansei Engineering, September 17, 2010*. (document)
- Ngiam, Jiquan, Khosla, Aditya, Kim, Mingyu, Nam, Juhan, Lee, Honglak, & Ng, Andrew Y. 2011. Multimodal deep learning. *Page 689–696 of: Proceedings of the 28th international conference on machine learning (ICML-11)*. (document)
- Picard, Rosalind Wright. 1995. Affective computing. (document)
- Rosenblatt, Frank. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, **65**(6), 386. (document)
- Salakhutdinov, Ruslan, Mnih, Andriy, & Hinton, Geoffrey. 2007. Restricted Boltzmann machines for collaborative filtering. *Page 791–798 of: Proceedings of the 24th international conference on Machine learning*. ACM. (document)
- Schuller, Björn, Steidl, Stefan, Batliner, Anton, *et al.* . 2009. The INTERSPEECH 2009 emotion challenge. *Page 312–315 of: INTERSPEECH*, vol. 2009. (document)
- Smolensky, Paul. 1986. *Information processing in dynamical systems: Foundations of harmony theory*. Tech. rept. DTIC Document. (document)
- Strapparava, Carlo, Valitutti, Alessandro, *et al.* . 2004. WordNet Affect: an Affective Extension of WordNet. *Page 1083–1086 of: LREC*, vol. 4. Citeseer. (document)
- Stuhlsatz, André, Meyer, Christine, Eyben, Florian, Zielke, Thomas, Meier, Günter, & Schuller, Björn. 2011. Deep neural networks for acoustic emotion recognition: raising the benchmarks. *Page 5688–5691 of: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. (document)
- Viola, Paul, & Jones, Michael. 2001. Rapid object detection using a boosted cascade of simple features. *Page I–I of: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE. (document)

Zeiler, Matthew D. 2012. ADADELTA: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*. (document)

# Lampiran A

## Log Keluaran

### MNIST

Log keluaran pada data MNIST dengan menggunakan CNN adalah sebagai berikut:

```
Start training
[TRAINING DATA] shapes: (60000, 1, 28, 28)
[TEST DATA] shapes: (10000, 1, 28, 28)
[TRAINING] Total epochs: 2
```

| Epoch # | Train err | Valid err | Time     |
|---------|-----------|-----------|----------|
| 1       | 0.103     | 0.04052   | 00:11:18 |
| 2       | 0.02805   | 0.02922   | 00:11:17 |

```
Start testing
precision recall f1-score support:
0 0.99 0.99 0.99 1006
1 0.99 1.00 1.00 1157
2 1.00 0.98 0.99 1019
3 0.99 0.99 0.99 1008
4 0.99 0.99 0.99 930
5 0.99 0.99 0.99 877
6 0.99 0.99 0.99 997
7 0.98 0.99 0.99 1040
8 0.97 0.99 0.98 940
9 0.99 0.98 0.99 1026
avg / total:
0.99 0.99 0.99 10000
```

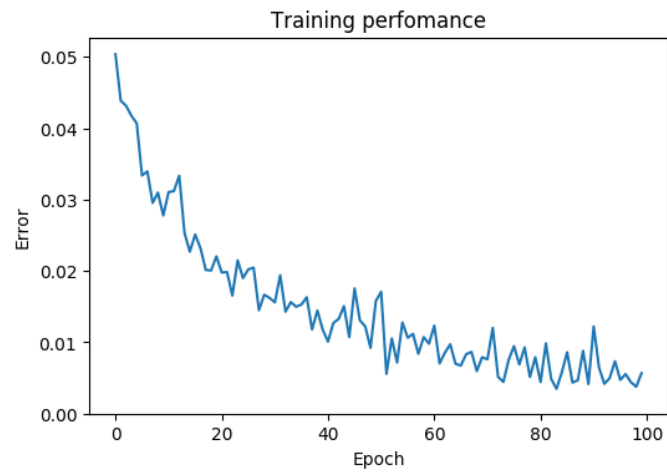
Validation accuracy: 98.98%

Gambar A.1: Keluaran log CNN pada data MNIST

Log keluaran pada data MNIST dengan menggunakan CSAE adalah sebagai berikut:

```
Start training
```

Gambar A.2: Keluaran log CSAE pada data MNIST



Gambar A.3: Grafik galat terhadap epoch CSAE pada MNIST

Log keluaran pada data MNIST dengan menggunakan CDBN adalah sebagai berikut:

Start training

Gambar A.4: Keluaran log CDBN pada data MNIST

.