

***Email Spam Filtering Menggunakan Multilayer Perceptron dengan  
Metode Pelatihan Stochastic Gradient Descent (SGD) dan  
Momentum***

*Diajukan Sebagai Syarat Untuk Menyelesaikan Pendidikan Program Strata-1  
Pada Jurusan Teknik Informatika*



Oleh:

Mgs. M. Luthfi Ramadhan  
09021381621077

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2020**

## LEMBAR PENGESAHAN TUGAS AKHIR

*Email Spam Filtering Menggunakan Multilayer Perceptron dengan Metode Pelatihan Stochastic Gradient Descent (SGD) dan Momentum*

Oleh :

Mgs. M. Luthfi Ramadhan  
NIM : 09021381621077

Palembang, Maret 2020

Mengetahui,  
a.n. Ketua Jurusan  
Sekretaris Jurusan



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

Pembimbing,



Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 21 Februari 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Mgs. M. Luthfi Ramadhan

NIM : 09021381621077

Judul : *Email Spam Filtering Menggunakan Multilayer Perceptron dengan Metode Pelatihan Stochastic Gradient Descent (SGD) dan Momentum*

1. Pembimbing

Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003



2. Penguji I

Dr. Abdiansah, S.Kom., M.Cs.  
NIP. 198410012009121005



3. Penguji II

Danny Matthew Saputra S.T., M.Sc.  
NIP. 198505102015041002



Mengetahui,  
a.n. Ketua Jurusan  
Sekretaris Jurusan



Alvi Svahrini Utami, M.Kom.  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Mgs. M. Luthfi Ramadhan  
NIM : 09021381621077  
Program Studi : Teknik Informatika  
Judul Skripsi : *Email Spam Filtering Menggunakan Multilayer  
Perceptron dengan Metode Pelatihan Stochastic Gradient  
Descent (SGD) dan Momentum*  
Hasil Pengecekan Software *iThenticate/Turnitin* : 7%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Maret 2020



Mgs. M. Luthfi Ramadhan  
NIM. 09021381621077

## MOTTO DAN PERSEMBAHAN

### MOTTO:

*“When Allah loves a servant, he test him”*

- Nabi Muhammad SAW

*“If Allah helps you, no one can defeat you”*

- Ali ‘Imran : 160

*“Being the underdog lit up the fire inside me”*

- Penulis

*Kupersembahkan karya tulis ini kepada:*

- ✓ *Allah SWT & Nabi Muhammad SAW*
- ✓ *Ibu & ayahku tercinta*
- ✓ *Keluarga besarku*
- ✓ *Dosen pembimbing & penguji*
- ✓ *Teman-temanku*
- ✓ *Almamater*

## **KATA PENGANTAR**

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Ibuku tercinta yang selalu mendoakan serta memberikan dukungan baik moril maupun materil.
2. Ayahku yang selalu menjadi motivasi dalam menjalani kehidupan.
3. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika, dan ibu Alvi Syahrini, M.T selaku Sekretaris Jurusan Teknik Informatika.
4. Bapak Samsuryadi, M.Kom.,Ph.D selaku dosen pembimbing yang telah membimbing, mengarahkan, dan memberikan motivasi dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak Dr. Abdiansah, S.Kom., M.Cs selaku dosen penguji I dan Bapak Danny Matthew Saputra, S.T., M.Sc selaku dosen penguji II yang telah memberikan masukan dan ilmu pengetahuan.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Mbak Wiwin dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.

8. Sahabat seperjuangan; Rosdiana, Melvin, Adi, dan Sari yang telah banyak mendukung dan memberi semangat selama proses penulisan tugas akhir ini.
9. Teman-teman yang tak henti-hentinya menghibur; Satria, Ahmad Gustano, Reza, Adi, dan Ade.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Maret 2020



Mgs. M. Luthfi Ramadhan

**Email Spam Filtering Using Multilayer Perceptron Trained with  
Stochastic Gradient Descent (SGD) and Momentum**

By:  
Mgs. M. Luthfi Ramadhan  
09021381621077

**ABSTRACT**


One of email spam is due to many advertisement has been sent by email. This phenomenon raised a very annoying problem which spend our time and money. This problem can be solved by developing email spam filter software with the help of artificial intelligence. Email spam filter make use of Multilayer Perceptron (MLP) trained with Stochastic Gradient Descent (SGD) and Momentum. 9 number of MLP's network architecture and 9 number of beta value are choosen. All of hidden layer's node use ReLU (Rectified Linear Unit) activation function, while output layer's node use Sigmoid activation funtion. Weights are uniformly randomized with range -1 until 1. The result of this final project show that architecture with 1 hidden layer which consist of 15 node and beta value of 0.8 produces higher F-measure.

Keyword: Email, Multilayer Perceptron, Stochastic Gradient Descent, Momentum.

Palembang, March 2020

Approved,  
o.b Head of Department  
Secretary of Department

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

Supervisor,  
  
Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003



*Email Spam Filtering Menggunakan Multilayer Perceptron dengan Metode Pelatihan Stochastic Gradient Descent (SGD) dan Momentum*

Oleh:  
Mgs. M. Luthfi Ramadhan  
09021381621077


**ABSTRAK**

Salah satu *email* spam sering disebabkan oleh banyaknya iklan yang dikirim melalui *email*. Fenomena ini menimbulkan masalah mengganggu yang menghabiskan waktu dan uang. Permasalahan ini dapat diselesaikan dengan mengembangkan perangkat lunak penyaring *email* spam dengan bantuan kecerdasan buatan. Penyaringan *email* spam menggunakan metode *Multilayer Perceptron* (MLP) dengan pelatihan *Stochastic Gradient Descent* (SGD) dan Momentum. Arsitektur jaringan MLP dan nilai *beta* pada momentum dibuat sebanyak 9 arsitektur dan 9 nilai *beta*. Semua lapis tersembunyi menggunakan fungsi aktivasi ReLU (*Rectified Linear Unit*), sedangkan lapis keluaran menggunakan fungsi aktivasi *Sigmoid*. Bobot diinisialisasi secara *random uniform* dengan *range* -1 sampai 1. Hasil penelitian ini menunjukkan bahwa arsitektur dengan 1 lapisan tersembunyi yang terdiri dari 15 neuron dan nilai *beta* 0,8 menghasilkan *f-measure* yang lebih tinggi.

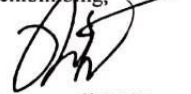
Kata kunci: *Email spam, Multilayer Perceptron, Stochastic Gradient Descent, Momentum.*

Palembang, Maret 2020

Mengetahui,  
a.n. Ketua Jurusan  
Sekretaris Jurusan

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

Pembimbing,

  
Samsuryadi, M.Kom., Ph.D.  
NIP. 197102041997021003

## DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN TUGAS AKHIR .....	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR .....	iii
HALAMAN PERNYATAAN .....	iv
MOTTO DAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
ABSTRACT .....	viii
ABSTRAK .....	ix
DAFTAR ISI .....	x
DAFTAR TABEL .....	xiv
DAFTAR GAMBAR .....	xv
BAB I PENDAHULUAN .....	I-1
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang.....	I-1
1.3    Rumusan Masalah .....	I-4
1.4    Tujuan Penelitian.....	I-5
1.5    Manfaat Penelitian.....	I-5
1.6    Batasan Masalah.....	I-5
1.7    Sistematika Penulisan.....	I-6
1.8    Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2 <i>Email</i> .....	II-1
2.3 <i>Text-Mining</i> .....	II-1
2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	II-2
2.5 <i>Multilayer Perceptron (MLP)</i> .....	II-3

2.6	Fungsi Aktivasi .....	II-4
2.6.1	<i>Sigmoid</i> .....	II-4
2.6.2	<i>Rectified Linear Unit (ReLU)</i> .....	II-4
2.7	<i>Stochastic Gradient Descent (SGD)</i> .....	II-5
2.8	Momentum .....	II-7
2.9	<i>Confusion Matrix</i> .....	II-7
2.9.1	<i>Recall (Sensitivity)</i> .....	II-8
2.9.2	<i>Precision</i> .....	II-9
2.9.3	<i>F-measure</i> .....	II-9
2.10	<i>Unified Modeling Language (UML)</i> .....	II-10
2.11	<i>Rational Unified Process (RUP)</i> .....	II-11
2.12	Penelitian Lain Yang Relevan .....	II-12
2.13	Kesimpulan.....	II-14
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data.....	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.2.2	Metode Pengumpulan Data.....	III-2
3.2.3	Pembagian <i>Dataset</i> .....	III-2
3.3	Tahapan Penelitian .....	III-3
3.3.1	Kerangka Kerja .....	III-5
3.3.2	Kriteria Pengujian .....	III-6
3.3.3	Format Data Pengujian .....	III-7
3.3.4	Alat yang digunakan dalam Pelaksanaan Penelitian .....	III-7
3.3.5	Pengujian Penelitian.....	III-8
3.3.6	Analisis Hasil Pengujian dan Penarikan Kesimpulan .....	III-8
3.4	Metode Pengembangan Perangkat Lunak .....	III-8
3.4.1	Fase Insepsi .....	III-9
3.4.2	Fase Elaborasi .....	III-9

3.4.3	Fase Konstruksi.....	III-10
3.4.4	Fase Transisi .....	III-10
3.5	Manajemen Proyek Penelitian.....	III-11

#### BAB IV PENGEMBANGAN PERANGKAT LUNAK..... IV-1

4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis.....	IV-1
4.2.2	Kebutuhan .....	IV-1
4.2.3	Analisis dan Perancangan.....	IV-2
4.2.4	Implementasi .....	IV-2
4.3	Fase Elaborasi.....	IV-3
4.3.1	Pemodelan Bisnis .....	IV-3
4.3.2	Kebutuhan .....	IV-13
4.3.3	Analisis dan Perancangan.....	IV-13
4.3.4	Implementasi .....	IV-34
4.4	Fase Konstruksi .....	IV-34
4.4.1	Pemodelan Bisnis .....	IV-34
4.4.2	Kebutuhan .....	IV-34
4.4.3	Analisis dan Perancangan.....	IV-35
4.4.4	Implementasi .....	IV-35
4.5	Fase Transisi.....	IV-37
4.5.1	Pemodelan Bisnis .....	IV-37
4.5.2	Kebutuhan .....	IV-38
4.5.3	Analisis dan Perancangan.....	IV-38
4.5.4	Implementasi .....	IV-41
4.6	Kesimpulan.....	IV-48

BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1    Pendahuluan .....	V-1
5.2    Hasil Percobaan Penelitian .....	V-1
5.2.1    Skenario Percobaan .....	V-1
5.2.1.1    Skenario Percobaan I .....	V-3
5.2.1.2    Skenario Percobaan II .....	V-4
5.2.1.3    Skenario Percobaan III.....	V-5
5.2.1.4    Skenario Percobaan IV.....	V-6
5.2.1.5    Skenario Percobaan V .....	V-7
5.2.1.6    Skenario Percobaan VI.....	V-8
5.2.1.7    Skenario Percobaan VII .....	V-9
5.2.1.8    Skenario Percobaan VIII.....	V-10
5.2.1.9    Skenario Percobaan IX.....	V-11
5.2.2    Perbandingan SGD dan SGD-Momentum.....	V-12
5.3    Analisis Hasil Penelitian.....	V-13
5.4    Kesimpulan.....	V-15
 BAB VI KESIMPULAN DAN SARAN .....	VI-1
6.1    Kesimpulan.....	VI-1
6.2    Saran.....	VI-1
 DAFTAR PUSTAKA .....	VII-1
 LAMPIRAN .....	L-1

## DAFTAR TABEL

Halaman

<b>Tabel III-1</b> Rancangan Hasil Pengujian.....	III-7
<b>Tabel III-2</b> Tabel <i>Work Breakdown Structure</i> (WBS) Penelitian.....	III-12
<b>Tabel IV-1</b> Kebutuhan Fungsional Perangkat Lunak.....	IV-2
<b>Tabel IV-2</b> Kebutuhan Non Fungsional Perangkat Lunak.....	IV-2
<b>Tabel IV-3</b> Definisi Aktor .....	IV-4
<b>Tabel IV-4</b> Definisi <i>Use-Case</i> .....	IV-4
<b>Tabel IV-5</b> Skenario <i>Use Case</i> Membuat model .....	IV-5
<b>Tabel IV-6</b> Skenario <i>Use Case</i> Save model .....	IV-6
<b>Tabel IV-7</b> Skenario <i>Use Case</i> Load model.....	IV-7
<b>Tabel IV-8</b> Skenario <i>Use Case</i> Clear model.....	IV-9
<b>Tabel IV-9</b> Skenario <i>Use Case</i> Mengevaluasi model .....	IV-10
<b>Tabel IV-10</b> Skenario <i>Use Case</i> Menyaring <i>email</i> .....	IV-11
<b>Tabel IV-11</b> Daftar Implementasi Kelas .....	IV-36
<b>Tabel IV-12</b> Rencana Pengujian <i>Use Case</i> Membuat model.....	IV-38
<b>Tabel IV-13</b> Rencana Pengujian <i>Use Case</i> Save model.....	IV-38
<b>Tabel IV-14</b> Rencana Pengujian <i>Use Case</i> Load model.....	IV-39
<b>Tabel IV-15</b> Rencana Pengujian <i>Use Case</i> Clear model.....	IV-39
<b>Tabel IV-16</b> Rencana Pengujian <i>Use Case</i> Mengevaluasi model.....	IV-39
<b>Tabel IV-17</b> Rencana Pengujian <i>Use Case</i> Menyaring <i>email</i> .....	IV-40
<b>Tabel IV-18</b> Pengujian <i>Use Case</i> Membuat model.....	IV-41
<b>Tabel IV-19</b> Pengujian <i>Use case</i> Save model .....	IV-43
<b>Tabel IV-20</b> Pengujian <i>Use Case</i> Load model.....	IV-44
<b>Tabel IV-21</b> Pengujian <i>Use Case</i> Clear model.....	IV-45
<b>Tabel IV-22</b> Pengujian <i>Use Case</i> Mengevaluasi model.....	IV-46
<b>Tabel IV-23</b> Pengujian <i>Use Case</i> Menyaring <i>email</i> .....	IV-47
<b>Tabel V-1</b> Skenario Percobaan .....	V-2
<b>Tabel V-2</b> Tabel Hasil Skenario Percobaan I .....	V-3
<b>Tabel V-3</b> Tabel Hasil Skenario Percobaan II.....	V-4
<b>Tabel V-4</b> Tabel Hasil Skenario Percobaan III .....	V-5
<b>Tabel V-5</b> Tabel Hasil Skenario Percobaan IV .....	V-6
<b>Tabel V-6</b> Tabel Hasil Skenario Percobaan V .....	V-7
<b>Tabel V-7</b> Tabel Hasil Skenario Percobaan VI .....	V-8
<b>Tabel V-8</b> Tabel Hasil Skenario Percobaan VII.....	V-9
<b>Tabel V-9</b> Tabel Hasil Skenario Percobaan VIII .....	V-10
<b>Tabel V-10</b> Tabel Hasil Skenario Percobaan IX .....	V-11
<b>Tabel V-11</b> Tabel Hasil Perbandingan SGD dan SGD-Momentum .....	V-12

## DAFTAR GAMBAR

Halaman

<b>Gambar II-1</b> Jaringan <i>Multilayer Perceptron</i> . .....	II-3
<b>Gambar II-2</b> <i>Confusion Matrix</i> .....	II-7
<b>Gambar II-3</b> Diagram Proses <i>Rational Unified Process (RUP)</i> .....	II-12
<b>Gambar III-1</b> Kerangka kerja klasifikasi <i>email</i> spam menggunakan MLP .	III-5
<b>Gambar IV-1</b> <i>Use case</i> Diagram .....	IV-3
<b>Gambar IV-2</b> Diagram Kelas Analisis Membuat model.....	IV-13
<b>Gambar IV-3</b> Diagram Kelas Analisis <i>Save</i> model.....	IV-14
<b>Gambar IV-4</b> Diagram Kelas Analisis <i>Load</i> model .....	IV-14
<b>Gambar IV-5</b> Diagram Kelas Analisis <i>Clear</i> model.....	IV-15
<b>Gambar IV-6</b> Diagram Kelas Analisis Mengevaluasi model.....	IV-15
<b>Gambar IV-7</b> Diagram Kelas Analisis Menyaring <i>email</i> .....	IV-16
<b>Gambar IV-8</b> <i>Activity</i> diagram membuat model .....	IV-17
<b>Gambar IV-9</b> <i>Activity</i> diagram <i>Save</i> model.....	IV-18
<b>Gambar IV-10</b> <i>Activity</i> diagram <i>Load</i> model.....	IV-19
<b>Gambar IV-11</b> <i>Activity</i> diagram <i>Clear</i> model .....	IV-20
<b>Gambar IV-12</b> <i>Activity</i> diagram Mengevaluasi model.....	IV-21
<b>Gambar IV-13</b> <i>Activity</i> diagram Menyaring <i>email</i> .....	IV-22
<b>Gambar IV-14</b> <i>Sequence</i> diagram membuat model .....	IV-23
<b>Gambar IV-15</b> <i>Sequence</i> diagram __textclean ( <i>Subsequence</i> ).....	IV-24
<b>Gambar IV-16</b> <i>Sequence</i> diagram __forward ( <i>Subsequence</i> ) .....	IV-25
<b>Gambar IV-17</b> <i>Sequence</i> diagram __Sgd_momentum ( <i>Subsequence</i> ).....	IV-26
<b>Gambar IV-18</b> <i>Sequence</i> diagram <i>Save</i> model.....	IV-27
<b>Gambar IV-19</b> <i>Sequence</i> diagram <i>Load</i> model .....	IV-28
<b>Gambar IV-20</b> <i>Sequence</i> diagram <i>Clear</i> model.....	IV-29
<b>Gambar IV-21</b> <i>Sequence</i> diagram Mengevaluasi model.....	IV-30
<b>Gambar IV-22</b> <i>Sequence</i> diagram <i>confuse</i> ( <i>Subsequence</i> ).....	IV-31
<b>Gambar IV-23</b> <i>Sequence</i> diagram Menyaring <i>email</i> .....	IV-32
<b>Gambar IV-24</b> Diagram Kelas .....	IV-33
<b>Gambar IV-25</b> Antarmuka perangkat lunak.....	IV-35
<b>Gambar IV-26</b> Tampilan antarmuka perangkat lunak.....	IV-36
<b>Gambar V-1</b> Arsitektur MLP skenario I.....	V-2
<b>Gambar V-2</b> Grafik Hasil Skenario Percobaan I.....	V-3
<b>Gambar V-3</b> Grafik Hasil Skenario Percobaan II.....	V-4
<b>Gambar V-4</b> Grafik Hasil Skenario Percobaan III .....	V-5
<b>Gambar V-5</b> Grafik Hasil Skenario Percobaan IV .....	V-6
<b>Gambar V-6</b> Grafik Hasil Skenario Percobaan V .....	V-7
<b>Gambar V-7</b> Grafik Hasil Skenario Percobaan VI.....	V-8
<b>Gambar V-8</b> Grafik Hasil Skenario Percobaan VII.....	V-9

<b>Gambar V-9</b> Grafik Hasil Skenario Percobaan VIII .....	V-10
<b>Gambar V-10</b> Grafik Hasil Skenario Percobaan IX .....	V-11
<b>Gambar V-11</b> Grafik Hasil Perbandingan SGD dan SGD-Momentum .....	V-12
<b>Gambar V-12</b> Grafik Selisih <i>F-measure</i> Antara Data Latih dan Data Uji....	V-13
<b>Gambar V-13</b> Lokal Minimum.....	V-15



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Pada bab pendahuluan akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Bab ini juga berisikan penjelasan mengenai gambaran umum dari keseluruhan kegiatan yang dilakukan dalam penelitian tugas akhir.

### **1.2 Latar Belakang**

Semakin banyaknya orang terhubung dengan internet, *electronical mail* (*email*) menjadi sarana komunikasi jarak jauh yang paling banyak digunakan karena murah, bermanfaat, cepat dan mudah untuk diakses (Roy dkk., 2013). *Email* tidak hanya digunakan untuk komunikasi, tetapi *email* juga digunakan sebagai media penyebaran iklan dalam bidang *electronic commerce* (*e-commerce*). Sebuah iklan yang dikirimkan kepada banyak orang melalui *email* tanpa persetujuan dari orang tersebut dinamakan *unsolicited commercial email* (UCE), *spam email*, *junk mail*, *bulk mail* atau *email* sampah.

Saat ini spam telah menjadi masalah serius bagi pengguna *email*. Spam sebenarnya bukanlah masalah yang merepotkan jika dalam jumlah yang sedikit. Tetapi sebaliknya, spam akan menjadi masalah yang merepotkan bagi pengguna jika masuk ke dalam kotak masuk (*inbox*) dalam jumlah yang banyak. Hal ini mengakibatkan banyaknya waktu terbuang hanya untuk membaca dan menghapus

*email* spam tersebut. Terlepas dari waktu yang terbuang untuk menghapus spam *mail* dari *inbox*, spam juga merupakan pemborosan uang bagi *user* yang menggunakan koneksi *dial-up* (Anugroho dkk., 2010).

Masalah ini dapat diselesaikan dengan membuat model klasifikasi yang mampu mengklasifikasikan *email* spam dengan akurasi tinggi. *Multilayer Perceptron* (MLP) adalah salah satu metode klasifikasi yang terkenal dengan kemampuannya dalam memahami pola data yang rumit. Setiap *design* dan arsitektur dari MLP akan menghasilkan hasil yang berbeda-beda. Hal ini sangat menarik perhatian para peneliti untuk melakukan penelitian mengenai MLP karena peneliti mendapatkan hasil memuaskan, seperti penelitian oleh Potghan dkk. (2018) menggunakan MLP dalam mendeteksi tumor paru-paru. Penelitian ini menggunakan 1000 lapis tersembunyi dan pelatihan dengan *lbfgs (Limited-memory Broyden Fletcher Goldfab Shanno)*. Hasil penelitian ini menunjukkan bahwa model MLP mampu melakukan deteksi tumor paru-paru dengan tingkat akurasi 98,31 %.

Rafii dkk. (2016) melakukan klasifikasi kanker paru-paru berdasarkan *microarray gene expression*. Pada penelitian ini dimensi direduksi menjadi 13 menggunakan PCA, dengan demikian maka lapis masukan terdiri dari 13 neuron. Hasil terbaik penelitian ini didapatkan dengan 7 neuron pada lapis tersembunyi pertama dan 4 neuron pada lapis tersembunyi kedua dengan akurasi 99,32 %.

Kaur & Prince Verma (2017) melakukan penelitian dengan judul “*E-mail Spam Detection Using Refined MLP with Feature Selection*”. Pada penelitian ini MLP dibandingkan dengan SVM dan pemilihan fitur menggunakan metode *N-gram*. Terdapat 6 skenario yaitu *Bi-gram SVM*, *Tri-gram SVM*, *Four-gram SVM*,

*Bi-gram* MLP, *Tri-gram* MLP, dan *Four-gram* MLP. Akurasi tertinggi adalah 81,62 % dengan skenario *Tri-gram* MLP. Meskipun MLP terkenal dengan kemampuannya dalam memahami pola data yang rumit, terdapat beberapa kekurangan dari MLP seperti proses pelatihan yang memakan waktu lama. Dikarenakan masalah tersebut, munculah beberapa metode optimasi seperti *Stochastic Gradient Descent* (SGD) dan Momentum yang dianggap dapat mempercepat proses pelatihan.

*Stochastic Gradient Descent* (SGD) adalah variasi dari *gradient descent* yang melakukan *update* pada setiap satu *sample* dari data latih. Ghofur (2018), melakukan penelitian mengenai MLP-SGD untuk peramalan nilai tukar mata uang Rupiah terhadap Dolar Amerika. Hasil penelitian menunjukkan nilai untuk RMSE, MAE, dan MAPE dengan masing-masing sebesar 61.992, 43.889, 0,33 dan Dstat 61,67%. Dengan demikian model ini telah memenuhi standar industri untuk model peramalan nilai tukar karena nilai Dstat sudah di atas 60 %. (Ghofur, 2018).

Abbas dkk. (2010), melakukan penelitian mengenai *Digit Recognition* dengan *Backpropagation* dan Momentum. Lapis masukan berjumlah 100 neuron, fungsi aktivasi lapis tersembunyi adalah sigmoid, dan lapis keluaran berjumlah 10 neuron. Hasil terbaik dari penelitian adalah 99,50 % didapatkan dengan 15 neuron pada lapis tersembunyi, 0,6 *beta* pada Momentum, dan 0,9 *learning rate* (Abbas dkk., 2010).

Berdasarkan latar belakang di atas. Tugas akhir ini melanjutkan penelitian Abbas dkk. (2010). Pada tugas akhir ini dilakukan penambahan SGD dan pengubahan kasus menjadi klasifikasi *email* spam.

### 1.3 Rumusan Masalah

Masalah *email* spam masih sering dialami pengguna *email*, ditambah maraknya bisnis-bisnis *online* yang memanfaatkan sarana komunikasi ini untuk tujuan iklan. Penelitian pada bidang jaringan syaraf tiruan telah banyak dilakukan, salah satunya pada kasus penyaringan *email* spam, dan salah satu metode yang sering digunakan adalah *Multilayer Perceptron* (MLP) yang dilatih dengan *Stochastic Gradient Descent* (SGD). Melatih MLP menggunakan SGD juga tidak selalu berjalan lancar dikarenakan SGD mudah terjebak di lokal minimum (sebuah permukaan seperti lembah namun bukan lembah yang paling rendah). Untuk mengatasi kekurangan tersebut, penelitian ini menggunakan metode momentum yang dianggap dapat membantu SGD keluar dari lokal minimum. Fokus permasalahan pada penelitian ini adalah mencari nilai *beta* dan arsitektur MLP yang dapat menghasilkan *f-measure* tinggi dalam klasifikasi *email* spam. Selanjutnya dirumuskan pertanyaan penelitian sebagai berikut:

1. Bagaimana mengembangkan arsitektur untuk klasifikasi *email* spam dengan *Multilayer Perceptron*?
2. Berapa nilai *beta* pada Momentum yang menghasilkan *f-measure* tinggi dalam klasifikasi *email* spam?

#### 1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Menghasilkan arsitektur *Multilayer Perceptron* untuk klasifikasi *email* spam.
2. Mengembangkan perangkat lunak untuk klasifikasi *email* spam dengan *Multilayer Perceptron*.
3. Mengetahui nilai *beta* pada Momentum yang menghasilkan *f-measure* tinggi dalam klasifikasi *email* spam.

#### 1.5 Manfaat Penelitian

Manfaat penelitian ini sebagai berikut:

1. Arsitektur klasifikasi *email* spam dapat digunakan sebagai referensi dan kajian yang berkaitan dengan klasifikasi *email* spam.
2. Perangkat lunak dapat digunakan untuk klasifikasi *email* spam.
3. Hasil penelitian dapat digunakan sebagai referensi dalam menentukan nilai *beta* pada momentum.

#### 1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. *Learning rate* yang digunakan adalah *learning rate* tetap dengan nilai 0,01.
2. Lapis tersembunyi menggunakan fungsi aktivasi ReLU.
3. Lapis keluaran terdiri dari satu neuron dengan *Sigmoid* sebagai fungsi aktivasi.

4. Evaluasi model dilakukan berdasarkan pada *f-measure* setelah 10 iterasi.
5. *Loss function* yang digunakan adalah *Binary Crossentropy*.
6. Fitur yang digunakan adalah 300 *term* yang paling sering muncul.
7. Perangkat lunak hanya mampu mengakses gmail.

### **1.7 Sistematika Penulisan**

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain:

### **BAB I. PENDAHULUAN**

Bab ini menguraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

### **BAB II. KAJIAN LITERATUR**

Bab ini membahas dasar-dasar teori yang akan digunakan dalam penelitian, seperti definisi *Email*, *Text-mining*, *Term Frequency-Inverse Document Frequency* (TF-IDF), *Multilayer Perceptron* (MLP), *Stochastic Gradient Descent* (SGD), Momentum, dan penelitian lain yang relevan.

### **BAB III. METODOLOGI PENELITIAN**

Bab ini akan membahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

## **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Bab ini akan membahas tahapan yang dilakukan dalam proses pengembangan perangkat lunak berdasarkan metode *Rational Unified Process* (RUP) yang mencakup fase insepisi, elaborasi, konstruksi, dan transisi.

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Bab ini akan membahas hasil pengujian dan analisis hasil pengujian.

## **BAB VI. KESIMPULAN DAN SARAN**

Bab ini berisikan kesimpulan dan saran berdasarkan pada hasil yang telah dibahas pada bab V.

### **1.8 Kesimpulan**

Pada bab 1 ini dapat disimpulkan bahwa masalah yang harus diselesaikan pada penelitian ini adalah berapa nilai *beta* pada momentum dan bagaimana arsitektur MLP agar dapat mengklasifikasikan *email* spam dengan *f-measure* tinggi.

## DAFTAR PUSTAKA

- Abbas, Q., Ahmad, J., & Bangyal, W. H. 2010. *Momentum term heals the performance of back propagation algorithm for digit recognition. Proceedings - 2010 6th International Conference on Emerging Technologies, ICET 2010*, 16–20. <https://doi.org/10.1109/ICET.2010.5638387>.
- Amrizal, V. 2018. Penerapan Metode *Term Frequency Inverse Document Frequency* (Tf-Idf) Dan *Cosine Similarity* Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim), *Jurnal Teknik Informatika*, 11(2), pp. 149–164. doi: 10.15408/jti.v11i2.8623.
- Anugroho, P., Winarno, I., & Rosyid, M. N. 2010. Klasifikasi *email* spam dengan metode *naïve bayes classifier* menggunakan *java programming*. *Journal Informatika*, 1–11.
- Awad, W. 2011. *Machine Learning Methods for Spam E-Mail Classification*. *International Journal of Computer Science and Information Technology*, 3(1), 173–184. <https://doi.org/10.5121/ijcsit.2011.3112>.
- Ghofur, M. M. 2018. *Stochastic Gradient Descent - Multilayer Perceptron* Untuk Peramalan Nilai Tukar Mata Uang Rupiah Terhadap Dolar Amerika. Skripsi Program Teknik Informatika FAKULTAS ILMU KOMPUTER Yogyakarta (tidak dipublikasikan).
- Goodfellow, I., Bengio, Y., & Courville, A. 2016. *Deep Learning*. MIT Press, London, England.



- Kaur, H., & Prince, V. E. 2017. E-Mail Spam Detection Using Refined MLP with Feature Selection. *International Journal of Modern Education and Computer Science*, 9(9), 42–52. <https://doi.org/10.5815/ijmecs.2017.09.05>.
- Langgeni, D. P. 2010 . ‘Komunitas Data *mining* Indonesia & *Soft-computing* Indonesia.’, *Text Mining dan Knowledge Discovery*., 0(semnasIF), pp. 1–10. doi: 10.1007/s13398-014-0173-7.2.
- Li, X. L. 2018. ‘*Preconditioned stochastic gradient descent*’, *IEEE Transactions on Neural Networks and Learning Systems*, 29(5), pp. 1454–1466. doi: 10.1109/TNNLS.2017.2672978.
- Pan, W., Narasimhan, H., Protopapas, P., Kar, P., & Ramaswamy, H. G. (2017). Optimizing the multiclass F-measure via biconcave programming. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1101–1106. <https://doi.org/10.1109/ICDM.2016.184>.
- Patterson, J. & Gibson A. 2017. *Deep Learning A Practitioner’s Approach*, O’Reilly Media, Sebastopol, USA.
- Potghan, S., Rajamenakshi, R., & Bhise, A. 2018. Multi-Layer Perceptron Based Lung Tumor Classification. *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018*, (Iceca), 499–502. <https://doi.org/10.1109/ICECA.2018.8474864>.

- Purwaningsih, N. 2016. Penerapan *Multilayer Perceptron* untuk klasifikasi Jenis Kulit Sapi Tersamak, *Jurnal TEKNOIF*, 4(1), pp. 1–7.
- Rafii, F., Kbir, M. A., & Hassani, B. D. R. 2016. *MLP network for lung cancer presence prediction based on microarray data. Proceedings of 2015 IEEE World Conference on Complex Systems, WCCS 2015.*  
<https://doi.org/10.1109/ICoCS.2015.7483276>.
- Rao, A. S., Avadhani, P., & Chaudhuri, N. B. 2016. *A Content-Based Spam E-Mail Filtering Approach Using Multilayer Perceptron Neural Networks, International Journal of Engineering Trends and Technology*, 41(1), pp. 44–45. doi: 10.14445/22315381/ijett-v41p210.
- Roy, S., Patra, A., Sau, S., Mandal, K., & Kunar, S. 2013. *An Efficient Spam Filtering Techniques for Email Account. American Journal of Engineering Research (AJER)*, 02(10), 63–73.
- Ruder, S. 2016. *An overview of gradient descent optimization algorithms.* 1–14.  
Retrieved from <http://arxiv.org/abs/1609.04747>.
- Sarigul, M., & Avci, M. 2017. *Performance comparison of different momentum techniques on deep reinforcement learning, Proceedings - 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2017*, pp. 302–306. doi: 10.1109/INISTA.2017.8001175.

Syahrian, F. 2016. Perbandingan Metode Optimasi *Stochastic Gradient Descent*, Adadelta, dan Adam pada Jaringan Saraf Tiruan dalam Klasifikasi Data Aritmia. Skripsi Program Teknik Informatika FAKULTAS ILMU KOMPUTER Yogyakarta (tidak dipublikasikan).

Yasin, A., & Abuhasan, A. 2016. *An Intelligent Classification Model for Phishing Email Detection*. International Journal of Network Security & Its Applications, 8(4), 55–72. <https://doi.org/10.5121/ijnsa.2016.8405>.