



SKRIPSI

KLASIFIKASI MULTILABEL INFORMASI *TWEET* BENCANA ALAM MENGGUNAKAN *CATEGORICAL BOOSTING* DAN OPTIMASI *BAYESIAN*

ENZELICA VICA CHRISTINA
NPM 21083010114

DOSEN PEMBIMBING

Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
Kartika Maulida Hindrayani S.Kom, M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

KLASIFIKASI MULTILABEL INFORMASI *TWEET* BENCANA ALAM MENGGUNAKAN *CATEGORICAL BOOSTING* DAN OPTIMASI *BAYESIAN*

ENZELICA VICA CHRISTINA
NPM 21083010114

DOSEN PEMBIMBING

Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
Kartika Maulida Hindrayani S.Kom, M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

KLASIFIKASI MULTILABEL INFORMASI *TWEET* BENCANA ALAM MENGGUNAKAN *CATEGORICAL BOOSTING* DAN OPTIMASI *BAYESIAN*

ENZELICA VICA CHRISTINA
NPM 21083010114

DOSEN PEMBIMBING

Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
Kartika Maulida Hindrayani S.Kom, M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**

LEMBAR PENGESAHAN

KLASIFIKASI MULTILABEL INFORMASI *TWEET* BENCANA ALAM MENGUNAKAN *CATEGORICAL BOOSTING* DAN OPTIMASI *BAYESIAN*

Oleh:
ENZELICA VICA CHRISTINA
NPM. 21083010114

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 10 September 2025:

Menyetujui,

Wahyu Syaifullah Jauharis Saputra, S.Kom.,
M.Kom.

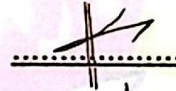
NIP. 19860825 202121 1 003



..... (Pembimbing I)

Kartika Maulida Hindrayani S.Kom, M.Kom.

NIP. 19920909 202203 2 009



..... (Pembimbing II)

Aviolla Terza Damaliana, S.Si., M.Stat.

NIP. 19940802 202203 2 015



..... (Ketua Penguji)

Shindi Shella May Wara, M. Stat.

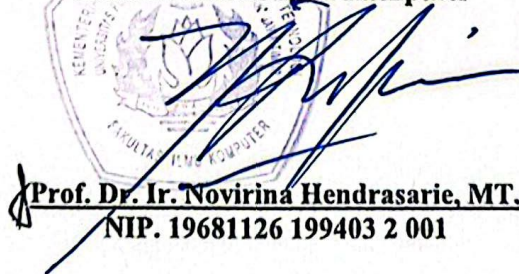
NIP. 19960518 202406 2 003



..... (Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

KLASIFIKASI MULTILABEL INFORMASI *TWEET* BENCANA ALAM MENGUNAKAN *CATEGORICAL BOOSTING* DAN OPTIMASI *BAYESIAN*

Oleh:
ENZELICA VICA CHRISTINA
NPM. 21083010114

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertandatangan di bawah ini:

Nama Mahasiswa : Enzelica Vica Christina
NPM : 21083010114
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya



Surabaya, 15 September 2025
Yang Membuat Pernyataan,



ENZELICA VICA CHRISTINA
NPM. 21083010114

ABSTRAK

Nama Mahasiswa / NPM : Enzelica Vica Christina / 21083010114
Judul Skripsi : Klasifikasi Multilabel Informasi *Tweet* Bencana Alam Menggunakan *Categorical Boosting* dan Optimasi *Bayesian*
Dosen Pembimbing : 1. Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
2. Kartika Maulida Hindrayani S.Kom, M.Kom.

Klasifikasi teks merupakan teknik penting dalam pemrosesan bahasa alami yang memungkinkan pengelompokan data teks secara otomatis. Twitter menjadi sumber data yang potensial untuk analisis kebencanaan karena bersifat *real time* serta banyak digunakan masyarakat Indonesia. Informasi yang terkandung di dalamnya sering mencakup lebih dari satu aspek, sehingga pendekatan klasifikasi multilabel lebih tepat dibandingkan klasifikasi tunggal. Namun, pemanfaatan algoritma *Categorical Boosting (CatBoost)* untuk klasifikasi multilabel data Twitter kebencanaan masih jarang dilakukan. Penelitian ini menerapkan *CatBoost* dengan keunggulan *Ordered Boosting* untuk mengurangi *overfitting* serta menangani data bising dan berdimensi tinggi. Performa model ditingkatkan melalui optimasi *Bayesian* yang mampu mengeksplorasi ruang parameter secara efisien pada kondisi distribusi label tidak seimbang. Model dikembangkan untuk memprediksi lima label informasi, yaitu verifikasi bencana, lokasi, kerusakan, korban, dan bantuan dengan evaluasi menggunakan *hamming loss*, *f1-score weighted*, dan *subset accuracy*. Kontribusi akademik penelitian ini terletak pada pembaruan metode, yaitu penggunaan *CatBoost* untuk klasifikasi multilabel teks, khususnya pada data Twitter kebencanaan yang masih jarang diteliti. Selain itu, kontribusi praktis diwujudkan melalui sistem berbasis *Streamlit* yang mampu menyajikan informasi bencana secara akurat dan efisien bagi kebutuhan tanggap darurat di Indonesia. Adapun hasil evaluasi menunjukkan bahwa skenario pembagian data 90:10 menghasilkan performa terbaik dengan *hamming loss* 0,0371, *f1-score weighted* 95,21%, dan *accuracy* 82,45%.

Kata kunci : Bencana Alam, *Categorical Boosting*, Klasifikasi Multilabel, Optimasi *Bayesian*, Twitter

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Enzelica Vica Christina / 21083010114
Thesis Title : *Multilabel Classification of Natural Disaster Tweet Information Using Categorical Boosting and Bayesian Optimization*
Advisor : 1. Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.
2. Kartika Maulida Hindrayani S.Kom, M.Kom.

ABSTRACT

Text classification is an important technique in natural language processing that enables automatic grouping of text data. Twitter is a potential source of data for disaster analysis because it is real time and widely used by Indonesians. The information contained therein often covers more than one aspect, so a multilabel classification approach is more appropriate than a single classification. However, the use of the Categorical Boosting (CatBoost) algorithm for multilabel classification of disaster Twitter data is still rare. This study applies CatBoost with the advantage of Ordered Boosting to reduce overfitting and handle noisy and high-dimensional data. Model performance is improved through Bayesian optimization, which is capable of efficiently exploring the parameter space under conditions of an imbalanced label distribution. The model was developed to predict six information labels, namely disaster, location, damage, victims, assistance, and others, with evaluation using hamming loss, weighted f1-score, and subset accuracy. Through this approach, the study provides an academic contribution in the form of a new reference for the application of CatBoost with Bayesian optimization for multilabel classification of disaster data. In addition, practical contributions are realized through a Streamlit-based system that is capable of presenting accurate and efficient disaster information for emergency response needs in Indonesia. The evaluation results show that the 90:10 data split scenario produces the best performance with a hamming loss of 0.0371, a weighted f1-score of 95.21%, and an accuracy of 82.45%.

Keywords: *Bayesian Optimization, Categorical Boosting, Multilabel Classification, Disaster Analysis, Twitter*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yesus Kristus atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **Klasifikasi Multilabel Informasi *Tweet* Bencana Alam Menggunakan *Categorical Boosting* dan Optimasi *Bayesian*** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom. selaku Dosen Pembimbing utama yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng. selaku Ketua Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran “ Jawa Timur.
3. Dosen-dosen Program Studi Sains Data yang telah memberikan ilmu dan wawasan selama perkuliahan maupun dalam proses penulisan skripsi ini.
4. Kedua orang tua tercinta, yang tidak pernah lelah memberikan kasih sayang, doa, dan dukungan tanpa henti. Setiap doa yang dipanjatkan menjadi sumber kekuatan, setiap nasihat menjadi penerang jalan, dan setiap pengorbanan menjadi motivasi terbesar penulis untuk tidak menyerah. Tanpa restu dan cinta yang tulus, penulis tidak akan mampu berdiri hingga titik ini.
5. Adik dan seluruh kerabat, yang senantiasa memberikan perhatian, semangat, dan doa. Kehadiran kalian membuat penulis merasa selalu ditemani dan tidak pernah sendirian dalam menapaki perjalanan panjang perkuliahan hingga penyusunan skripsi ini.
6. Teman-teman dekat, Zulfa, Lisa, Ica, Wenny, Elina, dan Mareta, yang telah menjadi bagian penting dalam perjalanan ini. Terima kasih atas kebersamaan, dukungan, tawa, serta cerita yang mengiringi setiap proses. Kehadiran kalian

bukan hanya sebagai teman, tetapi juga keluarga kedua yang membuat masa kuliah menjadi penuh warna.

7. Sahabat penulis, Chelia dan Ithin, yang meskipun terpisah jarak namun tetap hadir dengan dukungan, saran yang logis, dan usaha nyata dalam setiap langkah penulis. Terima kasih telah menjadi tempat berbagi, menguatkan ketika ragu, dan menunjukkan arti persahabatan sejati yang tidak pudar oleh waktu maupun jarak.
8. Diri penulis sendiri, yang telah berjuang melewati berbagai rintangan, keraguan, dan rasa lelah. Penulis menyadari bahwa perjalanan ini tidak selalu mudah; ada saat-saat penuh kegagalan, rasa ingin menyerah, bahkan kehilangan arah. Namun, penulis belajar untuk tetap berdiri, bangkit kembali, dan melanjutkan langkah meski dengan hati yang rapuh. Penulis berterima kasih kepada diri sendiri karena telah berani bermimpi, berusaha, dan tidak berhenti meskipun jalan yang ditempuh berliku. Semua proses ini menjadi bukti bahwa ketekunan, kesabaran, dan keyakinan dapat mengantarkan pada pencapaian yang dahulu hanya menjadi angan.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 15 September 2025

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI	xxi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	9
2.1. Penelitian Terdahulu	9
2.2. Dasar Teori.....	14
2.2.1. Bencana Alam	14
2.2.2. <i>Scraping</i>	15
2.2.3. <i>Text Preprocessing</i>	16
2.2.4. Ekstraksi Fitur (TF-IDF)	18
2.2.5. <i>Named Entity Recognition</i> (NER).....	19
2.2.6. Klasifikasi Multilabel.....	21
2.2.7. <i>Categorical Boosting</i>	22
2.2.8. Optimasi <i>Bayesian</i>	28
2.2.9. Evaluasi Model	30
2.2.10. <i>Streamlit</i>	32

BAB III DESAIN DAN IMPLEMENTASI SISTEM.....	33
3.1. Variabel Penelitian dan Sumber Data.....	33
3.1.1. Variabel Penelitian	33
3.1.2. Sumber Data	34
3.2. Langkah Analisis	35
3.2.1. Studi Literatur.....	35
3.2.2. Pengumpulan Data.....	36
3.2.3. <i>Preprocessing</i>	36
3.2.4. Pelabelan.....	37
3.2.5. Eksplorasi Data.....	38
3.2.6. Ekstraksi Fitur (<i>TF-IDF</i>)	38
3.2.7. Pemodelan dan Optimasi	39
3.2.8. Evaluasi Model.....	42
3.2.9. Implementasi <i>GUI</i>	43
3.3. Desain Sistem	43
BAB IV HASIL DAN PEMBAHASAN	47
4.1. Pengumpulan Data.....	47
4.2. <i>Preprocessing</i>	48
4.2.1. <i>Cleansing</i>	48
4.2.2. <i>Case Folding</i>	49
4.2.3. <i>Tokenizing</i>	50
4.2.4. Normalisasi Kata	51
4.2.5. <i>Stemming</i>	52
4.2.6. <i>Stopword Removal</i>	54
4.3. Pelabelan.....	55
4.3.1. Label Verifikasi Bencana	56
4.3.2. Label Lokasi	57
4.3.3. Label Kerusakan, Korban, dan Bantuan.....	59
4.4. Eksplorasi Data.....	63
4.5. Ekstraksi Fitur	66
4.6. Pemodelan dan Optimasi.....	69
4.6.1. Model <i>Categorical Boosting</i>	69

4.6.2.	Hasil Pemodelan <i>Categorical Boosting</i>	74
4.6.3.	Optimasi <i>Bayesian</i>	80
4.6.4.	Hasil Pemodelan Menggunakan Optimasi <i>Bayesian</i>	85
4.7.	Evaluasi	91
4.8.	Perbandingan Algoritma	93
4.9.	Implementasi <i>GUI</i>	94
4.9.1.	Beranda	95
4.9.2.	Halaman Data	96
4.9.3.	Halaman Persiapan Data	98
4.9.4.	Halaman Pemodelan	101
4.9.5.	Halaman Prediksi	104
BAB V	PENUTUP	107
5.1.	Kesimpulan	107
5.2.	Saran Pengembangan	108
DAFTAR PUSTAKA		111
LAMPIRAN		117

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja <i>Gradient Boosting</i>	22
Gambar 3.1 Diagram Alur Penelitian.....	35
Gambar 3.2 Alur Pemodelan dan Optimasi	39
Gambar 3.3 <i>Flowchart</i> Proses <i>CatBoost</i>	40
Gambar 3.4 <i>Flowchart</i> Proses Optimasi <i>Bayesian</i>	41
Gambar 3.5 <i>Flowchart</i> Sistem GUI	43
Gambar 3.6 Desain Halaman Beranda	44
Gambar 3.7 Desain Halaman <i>Upload Data</i>	44
Gambar 3.8 Desain Halaman Persiapan Data	45
Gambar 3.9 Desain Halaman Pemodelan	45
Gambar 3.10 Desain Halaman Klasifikasi Data Baru.....	46
Gambar 4.1 Diagram Distributi Label	63
Gambar 4.2 <i>Wordcloud</i> Komentar	64
Gambar 4.3 Kesalahan Klasifikasi <i>CatBoost</i> (90:10)	75
Gambar 4.4 Kesalahan Klasifikasi <i>CatBoost</i> (80:20)	77
Gambar 4.5 Kesalahan Klasifikasi <i>CatBoost</i> (70:30)	78
Gambar 4.6 Kesalahan Klasifikasi <i>CatBoost</i> (60:40)	79
Gambar 4.7 Kesalahan Klasifikasi <i>CatBoost</i> Setelah Optimasi (90:10).....	86
Gambar 4.8 Kesalahan Klasifikasi <i>CatBoost</i> Setelah Optimasi (80:20).....	88
Gambar 4.9 Kesalahan Klasifikasi <i>CatBoost</i> Setelah Optimasi (70:30).....	89
Gambar 4.10 Klasifikasi Kesalahan <i>CatBoost</i> Setelah Optimasi (60:40).....	90
Gambar 4.11 Tampilan Halaman Beranda.....	96
Gambar 4.12 Tampilan Upload Data	97
Gambar 4.13 Tampilan Awal Halaman Pra-proses.....	99
Gambar 4.14 Tampilan Preprocessing	100
Gambar 4.15 Tampilan Pelabelan Otomatis	100
Gambar 4.16 Tampilan TF-IDF	101
Gambar 4.17 Tampilan Halaman Pemodelan (I)	103
Gambar 4.18 Tampilan Halaman Pemodelan (II)	104
Gambar 4.19 Tampilan Halaman Prediksi	106

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 2.2 Contoh <i>Cleansing – Case Folding</i>	16
Tabel 2.3 Contoh <i>Tokenizing – Noemalisasi Kata</i>	17
Tabel 2.4 Contoh <i>Stemming – Stopword Removal</i>	18
Tabel 3.1 Variabel Penelitian	33
Tabel 3.2 Contoh Data Komentar Twitter	34
Tabel 3.3 Contoh Pelabelan	38
Tabel 4.1 Data Komentar	47
Tabel 4.2 Hasil Proses <i>Cleansing</i>	49
Tabel 4.3 Hasil Proses <i>Case Folding</i>	50
Tabel 4.4 Hasil Proses <i>Tokenizing</i>	51
Tabel 4.5 Hasil Proses Normalisasi Kata	52
Tabel 4.6 Hasil Proses <i>Stemming</i>	53
Tabel 4.7 Hasil Proses <i>Stopword Removal</i>	55
Tabel 4.8 Contoh Hasil Pelabelan	62
Tabel 4.9 Kombinasi Label Beserta Jumlahnya	66
Tabel 4.10 Contoh Data	67
Tabel 4.11 Contoh Perhitungan TF-IDF	68
Tabel 4.12 Parameter untuk Melatih Model	69
Tabel 4.13 Contoh Data dan Label	71
Tabel 4.14 Label Lokasi	72
Tabel 4.15 Pembaruan dan Pada Label Lokasi	74
Tabel 4.16 Evaluasi Global Rasio 90:10	75
Tabel 4.17 Evaluasi Global Rasio 80:20	76
Tabel 4.18 Evaluasi Global Rasio 70:30	77
Tabel 4.19 Evaluasi Global Rasio 60:40	79
Tabel 4.20 Ruang Pencarian Parameter Optimal	80
Tabel 4.21 Kombinasi Parameter Hasil Optimasi	82
Tabel 4.22 Evaluasi Global Setelah Optimasi (90:10)	86
Tabel 4.23 Evaluasi Global Setelah Optimasi (80:20)	87

Tabel 4.24 Evaluasi Global Setelah Optimasi (70:30)	88
Tabel 4.25 Evaluasi Global Setelah Optimasi (60:40)	90
Tabel 4.26 Perbandingan Performa Model.....	91
Tabel 4.27 Hasil Prediksi dengan Model Paling Optimal	93
Tabel 4.28 Perbandingan Algoritma.....	94

DAFTAR LAMPIRAN

Lampiran 1. LoA Jurnal Sinta 3	117
Lampiran 2. Data Penelitian.....	118
Lampiran 3. <i>Source Code CatBoost</i> dan Optimasi <i>Bayesian</i>	119
Lampiran 4. Kamus <i>Preprocessing</i>	119
Lampiran 5. Kamus Pelabelan	120
Lampiran 6. <i>Source Code</i> GUI.....	121

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

$Tf_{t,d}$	: jumlah <i>term</i> pada dokumen
df_t	: jumlah dokumen yang mengandung <i>term</i>
N	: jumlah keseluruhan dokumen
$F_m(X) = F_{m-1}(X)$	: model ke- dan pada iterasi
α_i	: parameter learning rate
r_{m-1}	: residual dari model sebelumnya
h_i	: model residual ke-
X_k	: input model <i>CatBoost</i>
y_k	: nilai target
I	: jumlah iterasi
σ	: permutasi acak dari indeks 1, ...,
M_i	: model prediksi untuk langkah ke-
r_i	: residual data ke-
y_i	: nilai target pada data ke-
$M_{\sigma(i)-1}(x_i)$	: prediksi saat ini
ΔM	: model tambahan
(x_j, r_j)	: data tersedia
$\hat{x}_k^i$	: Nilai fitur kategorikal ke- pada data ke- setelah diubah menggunakan OTS
D_k	: Himpunan data sebelumnya dari urutan acak hingga data ke-
$1_{x_j^i=x_k^i}$	: Fungsi indikator, bernilai 1 jika nilai kategorikal ke- dari data ke- sama dengan data ke- dan 0 jika tidak
y_j	: Label target dari data ke-
p	: Prior, biasanya nilai rata-rata dari target di seluruh data
a	: Parameter <i>smoothing</i>
$\mu_t(\mathbf{x})$	: Nilai ekspektasi (rata-rata prediksi)
$\sigma_t(\mathbf{x})$	: Ketidakpastian epistemik

$f(\mathbf{x})$	: Fungsi acak
$m(\mathbf{x})$	: Fungsi rata-rata, yaitu nilai ekspektasi dari f di titik $\mathbf{x}$
$k(\mathbf{x}, \mathbf{x}')$	: Fungsi kernel
$\mathbf{x}$ dan $\mathbf{x}'$	: Dua titik input dalam ruang desain
$\ \mathbf{x} - \mathbf{x}'\ ^2$	: Jarak <i>euclidean</i> kuadrat anatar $\mathbf{x}$ dan $\mathbf{x}'$
θ	: Parameter skala panjang (<i>length-scale</i>)
$\mathbf{k}$	: Vektor kovarian antar titik $\mathbf{x}$ terhadap semua titik observasi sebelumnya
$\mathbf{K}$	: Matriks kovarian dari semua titik observasi sebelumnya
σ_{noise}^2	: Variansi <i>noise</i> (<i>error</i> pengukuran)
$\mathbf{I}$	: Matriks identitas
$y_{1:t}$	: Nilai fungsi objektif pada titik $\mathbf{x}_1$ hingga $\mathbf{x}_t$
p	: Jumlah data yang diklasifikasikan
Q	: Jumlah total label
$ h(\mathbf{x}_i) - Y_i $	: Jumlah kesalahan klasifikasi yang terjadi